Prognostic Models in Dengue

Thesis

UNRESTRICTED

Prognostic Models in Dengue

by

PHUNG KHANH LAM

This thesis was submitted to the Open University UK
for the degree of Doctor of Philosophy in Life Sciences

Oxford University Clinical Research Unit
Hospital for Tropical Diseases
Ho Chi Minh City, Vietnam

April 2015

Date of Submission: 9 February 2015
Date of Award: 16 March 2015

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Reliable prediction models in dengue would facilitate early identification of patients likely to progress to more severe disease, potentially improving patient management. However, most published studies have limitations with respect to their modelling strategy, sample size, and chosen clinical outcomes, and to date none have exploited longitudinal data. Moreover, only a few studies have examined outcomes in patients presenting with dengue shock syndrome (DSS), the most severe form of the disease.

This thesis aims to overcome these limitations by using two large prospective datasets describing a) 1719 children with established DSS and b) 2598 children hospitalized with dengue. **First**, the population of children with DSS was characterized, and profound DSS, a composite outcome reflecting the need for intensive supportive care, was established as a suitable outcome for prognostic research in this population. **Second**, risk factors for profound DSS were identified and included in a robust prediction model. Based on this model, a simple score chart for use in clinical practice was derived. **Third**, risk factors for progression to DSS among children hospitalized with dengue were identified, and a prognostic model for progression to DSS was carefully developed. However, this model displayed only moderate performance and had limited clinical utility. **Lastly**, differences between acute and chronic diseases, and the implications for dynamic prediction modeling based on longitudinal data, are discussed. A case study of dynamic prediction modeling for development of DSS suggested that (1) the current platelet count can be used to improve baseline models that rely on enrolment values only, and (2) simple conditional dynamic models displayed similar performance to more complex joint models in this situation.

Acknowledgments

The first and most gratefulness I would like to give to my primary supervisor, Dr Marcel Wolbers. His wisdom and knowledge inspired and motivated me to go through this challenging but interesting PhD program. I am sincerely thankful for his patience and encouragement, which not only allow me to finish this piece of work but will also enable me to go further in my future career.

I would like to express my gratitude to Dr Bridget Wills, my co-supervisor, for her invaluable support for my study. As a supervisor, she taught me how to do work seriously and carefully . As a clinician, she taught me about the important of a careful and comprehensive assessment of patient. These knowledge will no doubt help me a lot in my future career.

I would like to also give my thankfulness to Dr Cameron Simmons, Ms Thanh Kieu, Ms Hanh Tien, Dr Hoai Tam, Dr The Trung, other members of the Dengue group and the Hospital for Tropical Diseases for their invaluable support through my PhD. Without them, I would not have a great opportunity to carry out this work.

A thankfulness should also be given to all of my colleagues in the Biostatistics group: Mr Duc Anh, Dr Nhan, Dr Hoang Nhat. Their comments and suggestions provided me a chance to think more carefully about my problem and also to learn about great skills that they possess.

Last but not least, I am indebted to my beloved family, my mother, my wife and my little beautiful daughter for all of their sincere support during the years that I conducted my study. Without their smile and encouragement, I would not have enough power to overcome all obstacles and challenges during this long period.

Contents

List of Tables

List of Figures

Abbreviations

ADE	antibody-dependent enhancement.
AIC	Akaike information criterion.
ALT	alanine aminotransferase.
AST	aspartate aminotransferase.
AUC	area under the ROC curve.
BIC	Bayesian information criterion.
BP	blood pressure.
CART	classification and regression trees.
CI	confidence interval.
CRF	case report form.
CVP	central venous pressure.
DENV	dengue virus.
DF	dengue fever.
DHF	dengue haemorrhagic fever.
DSS	dengue shock syndrome.
ELISA	enzyme-linked immunosorbent assay.
GAM	generalized additive models.
GI	gastrointestinal bleeding.
HCT	haematocrit.
HDU	high-dependency unit.
HI	haemagglutination-inhibition.
HR	hazards ratio.
HTD	Hospital for Tropical Diseases.

ICU intensive care unit.

IgA immunoglobulin A.

IgG immunoglobulin G.

IgM immunoglobulin M.

IQR interquartile range.

IV intravenous.

JE Japanese encephalitis.

MCAR missing completely at random.

MICE multivariate imputation by chained equations.

MRI magnetic resonance imaging.

N/A not available.

NS not specified.

OFI other febrile illness.

OR odds ratio.

OUCRU Oxford University Clinical Research Unit.

PCR polymerase chain reaction.

PICU paediatric intensive care unit.

PLT platelet count.

PP pulse pressure.

PRNT plaque reduction neutralization test.

RCT randomised controlled trial.

RNA ribonucleic acid.

RT-PCR reverse transcriptase polymerase chain reaction.

WHO World Health Organization.

Chapter 1

Introduction and aims of the thesis

Summary

This chapter provides an overview of dengue, and describes risk factors for severe dengue and the current status of prognostic research in dengue. The chapter concludes by presenting the clinical and statistical aims of this PhD thesis.

1.1 Overview of dengue

1.1.1 Epidemiology

Dengue is one of the most important mosquito-borne viral infections that affects humans worldwide (World Health Organization, 2012b). A recent cartographic analysis using a compiled database of 8309 geo-located records of dengue incidence occurring from 1960 to 2012 provided an estimate of 390 (95% credible interval 284-528) million dengue infection per year for the global population size of 2010, of which 25% are apparent infections (Bhatt et al., 2013). The disease also affects a large geographical area including more than 100 countries in the 4 continents Africa, Asia, Americas and Oceania (Bhatt et al., 2013) with an on-going spread to previously unaffected areas (La Ruche et al., 2010; Gjenero-Margan et al., 2011; Pun, 2011).

In Vietnam, the first dengue outbreak with virological investigation was in the Mekong Delta region in 1963 (Halstead et al., 1965). Nowadays, dengue is highly endemic in Vietnam and it is considered as the most frequent cause of fever amongst subjects presenting to the public primary health services in southern Vietnam (Phuong et al., 2006). A recent estimate of the burden of the disease in this country is 2.6 (95% credible interval 1.9-3.6) million of apparent infections and 7.9 (95% credible interval 6.1-10.4) million of inapparent infections annually (Bhatt et al., 2013). According to the World Health Organization (WHO), Vietnam is ranked third amongst the 30 most highly endemic countries/territories (after Brazil and Indonesia) (World Health Organization, 2012b).

Dengue can affect susceptible people irrespective of age from infants to the elderly; but because of acquired host immunity in endemic countries, it is more prevalent and more severe in children. In Vietnam, the disease predominantly affected children from 5 to 14 years before 1998, but since then, the number of adult cases has increased (Quang Ha et al., 2000) which can be explained by the change in age structure of the population (Cummings et al., 2009; Cuong et al., 2011).

The disease transmission depends on many factors including host, mosquitoes, viruses as well as environment (Guzman and Harris, 2014). In areas where dengue is endemic, there is a strong seasonality with peak dengue incidence during the rainy season (Nisalak et al., 2003; Cuong et al., 2013), and evidence of spatial dependence within fine and

intermediate scales (Salje et al., 2012; Cuong et al., 2013). There is also evidence that the spatio-temporal transmission of dengue depends heavily on the local human movement (Stoddard et al., 2013).

Even though dengue infection is self-limiting in most cases, it is still a potentially fatal disease. According to one estimate from 2012, there are hundreds of thousands of cases with severe dengue occurring annually, including 20,000 deaths (World Health Organization, 2012a). The large case numbers indicate that dengue infection puts a huge burden on health care systems, especially in developing countries. A prospective study in eight countries in the Americas and Asia estimated the per-patient cost for ambulatory and hospitalized cases to be 514 USD and 1394 USD, respectively (Suaya et al., 2009). In Vietnam, a recent estimate of the annual health care cost of dengue infection is 30.3 million USD (Shepard et al., 2013). In Can Tho, a city in Southern Vietnam, the average cost for a patient with severe dengue infection was 2,798,000 VND ($\sim$ 168 USD) (Tam et al., 2012b).

1.1.2 Dengue virus

The dengue virus (DENV) is a single-stranded, positive-sense ribonucleic acid (RNA) virus and belongs to the genus Flavivirus, family Flaviviridae. Its genome encodes 10 proteins including 3 structural proteins (capsid protein C, premembrane protein prM, and envelope protein E) and 7 non-structural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5). Dengue virus might have evolved as an infection of non-human primates thousands years ago but now humans are the main host and Aedes mosquitoes (especially Aedes aegypti) are the principal vectors (Holmes and Twiddy, 2003).

Currently, there are 4 established virus serotypes (DENV-1, DENV-2, DENV-3, DENV-4) that co-circulate in many regions of the world (Messina et al., 2014). Dengue infection triggers long-lived serotype-specific immunity and short-lived cross-immunity between serotypes (Simmons et al., 2012a). According to the antibody-dependent enhancement (ADE) hypothesis, secondary infection with a different virus serotype could result in a more severe disease (Halstead and O'Rourke, 1977; Guzman et al., 2013). Within each serotype, there are multiple genotypes which may have different structures that may lead to different virulence (Leitmeyer et al., 1999). In endemic areas, there is a serotype-

specific dengue virus circulation with sequential replacement of the dominant serotype (Endy, 2002; Nisalak et al., 2003; Vu et al., 2010) which could be explained by the increase in susceptibility to secondary infection of cases with primary infection, the increase in transmissibility during secondary infection due to higher viraemia (Recker et al., 2009) and other extrinsic factors including changes in vector density, infection rate or environmental temperature (Nisalak et al., 2003).

1.1.3 Clinical manifestations

The typical evolution of dengue disease is characterized by 4 phases: incubation period, febrile illness, critical phase and recovery phase (Simmons et al., 2012a; World Health Organization, 2009). The incubation period is without clinical features and lasts around 6.5 days (range 2.6-14.2) (Snow et al., 2014). During the febrile illness phase which usually lasts for 3.2 days (range 0.2-6.8), the most common symptoms are leucopenia and rash (Snow et al., 2014). Other non-specific symptoms can occur including headache, vomiting, myalgia, and mild haemorrhage (petechiae or bruising). Other laboratory abnormalities also occur during this phase including thrombocytopenia and increases in hepatic transaminases. Around the day of defervescence (day 4 to day 7 of illness), a small proportion of patients progress into the critical phase which is characterized by plasma leakage and associated clinical symptoms including pleural effusion, ascites and hypovolaemic shock, and other important laboratory changes such as haemoconcentration and hypoproteinaemia. Thrombocytopenia and haemostatic dysfunction are invariably present in patients with significant plasma leakage, and may result in severe haemorrhage, primarily from the gastrointestinal tract, exacerbating the overall severity of the infection. Organ dysfunction may also occur, usually secondary to profound shock or severe haemorrhage, but is sometimes a primary problem, especially in adults (Moxon and Wills, 2008). The last phase is the recovery phase, which usually lasts for 48-72 hours, and is characterised by the re-absorption of extra-vascular fluid and a general recovery to a normal health status. In general, most patients with dengue infection recover completely and only a small proportion of cases would die or acquire complications (mostly related to dengue shock syndrome (DSS)).

1.1.4 Diagnosis and classifications

Dengue diagnostics

As dengue infection has a wide spectrum of manifestations with mostly non-specific symptoms, a diagnosis based solely on clinical symptoms is unreliable (World Health Organization, 2009). Therefore, a number of laboratory tests have been developed to enable more accurate and specific dengue diagnosis. In general, these tests either aim to directly identify the presence of the virus or its genome/antigens, or use indirect methods which identify the presence of the virus indirectly through detecting the presence of antibodies that the host produced to eliminate the infection (Peeling et al., 2010).

Direct methods These methods include virus isolation, RNA detection by polymerase chain reaction (PCR) assays, and antigen detection (Peeling et al., 2010; World Health Organization, 2009). These methods are highly specific to dengue infection but can only be used in the early phase of disease. Suitable specimens are whole blood, serum, plasma or host tissues, but all should be collected before day 5-6 of illness. Amongst them, antigen detection using NS1-based assays is the fastest and cheapest method. However, their sensitivity is quite low comparing to their very high specificity, and their diagnostic accuracy may depend on viraemia level and immunity responses (Hang et al., 2009; Tricou et al., 2010b; Da Costa et al., 2014; Zhang et al., 2014).

Indirect methods These methods include enzyme-linked immunosorbent assay (ELISA) to detect immunoglobulin M (IgM), immunoglobulin G (IgG), and immunoglobulin A (IgA) antibodies in whole blood, serum or plasma collected during the late phase of disease, haemagglutination-inhibition (HI) assay to determine level of antibodies, and the plaque reduction neutralization test (PRNT) to assess the level of protective antibodies (Peeling et al., 2010). ELISA tests are less specific but cheaper and easier to conduct than direct methods and other indirect methods (World Health Organization, 2009).

Differentiation between primary and secondary dengue infection

To differentiate between primary and secondary dengue infections, researchers can rely on the HI test, the dengue PRNT and more pragmatic approaches based on IgM and IgG

capture ELISA. The HI test is problematic for several reasons: it requires paired specimens with at least 7 days in between, yet second specimens are often unavailable; it is time–consuming and technically cumbersome as it requires extra procedures to eliminate non-specific haemagglutination in samples; it is not specific to dengue as the test is also reactive to other flaviviruses (World Health Organization, 2009; Cordeiro et al., 2009; Innis et al., 1989; Kuno et al., 1991; Vaughn et al., 1999; Shu et al., 2003b; De Souza et al., 2004; Matheus et al., 2005). Similarly, PRNTs are difficult to perform and require considerable time to get results (Cordeiro et al., 2009). Therefore, many research groups have used more pragmatic approaches utilizing the ratio between IgM and IgG, or IgG alone using just a single specimen. However, results from these approaches require careful interpretation as the level of IgG antibody or the IgM/IgG antibody ratio depend on the test that was used and the day of illness when the specimen was obtained.

Dengue classification

Based mainly on studies of dengue in children from Thailand from the 1960s, the WHO case classification for dengue infection was first published in 1975 and then updated in 1986 and 1997 (Bandyopadhyay et al., 2006; Halstead, 2013). This classification differentiates dengue infection into 2 distinct entities: dengue fever (DF) and dengue haemorrhagic fever (DHF) and then sub-categorizes the DHF group further into 4 ordinal grades of increasing severity (I, II, III/IV or DSS), based on non-specific signs and the presence of haemorrhagic tendency and plasma leakage (World Health Organization, 1997). Over the last 30 years, several shortcomings of this classification have been identified including difficulties in practically assessing strict criteria for DHF, the low sensitivity in identifying severe dengue, the low discrimination ability of certain criteria such as the tourniquet test, and an overlap of clinical manifestations between categorizations (Bandyopadhyay et al., 2006; Hadinegoro, 2012). Therefore, following a large multi-centre studies which re-assessed the utility of the current guideline, a global expert consensus meeting was set up in 2009 to propose and then implement a new dengue classification (Alexander et al., 2011; Horstick et al., 2012). This new classification categorizes dengue into 2 levels of severity: dengue with/without warning signs and severe dengue (including severe plasma leakage leading to DSS or fluid accumulation with respiratory distress; severe bleeding;

and severe organ failure), with allowance for the possibility that patients in the former group may progress to the latter group (World Health Organization, 2009). This classification is much simpler than its predecessor and aims is to improve case management rather than being purely a research tool (Farrar et al., 2013). While the new system has not been without its critics, mainly regarding the non-specificity of warning signs and the potential adverse impact on patho-physiological research (Kalayanarooj, 2011; Srikiatkhachorn et al., 2011; Halstead, 2013), current evidence from large multi-centre study shows that the new classification has a high applicability and is perceived as user-friendly (Barniol et al., 2011). In addition, a large multi-country study is ongoing to assess the predictive value of the suggested warning signs for severe dengue (Horstick et al., 2012; Jaenisch et al., 2013).

1.1.5 Treatment and Prevention

Currently, there is no specific therapy available for dengue other than supportive care (Simmons et al., 2012a). Recent attempts to treat dengue infection using immune modulation (Tam et al., 2012a) and anti-viral therapy (Tricou et al., 2010a; Nguyen et al., 2013; Low et al., 2014) did not show clear benefits. However, although disappointing, these trials have provided a structural framework for further clinical trials in this field (Simmons et al., 2012b); currently a study of Lovastatin, a drug with both antiviral and endothelial stabilizing properties is in progress in Vietnam (Whitehorn et al., 2012).

Supportive treatment, especially careful monitoring and appropriate usage of fluid replacement are still the basis for successful management (Simmons et al., 2012a; World Health Organization, 2012c). All patients with warning signs who are unable to tolerate oral fluids should be hospitalized for close observation. Patients with severe dengue require emergency treatment including fluid resuscitation for shock and/or blood transfusion for severe bleeding and/or other supportive care and adjuvant therapy (inotropic therapy, renal replacement therapy, etc.). Effective fluid resuscitation is very important in managing patients with DSS. As the severity of leakage in patients with DSS ranges from relatively mild to severe plasma leakage, fluid requirements can range from a small prompt initial volume resuscitation to the need for very large volumes of parenteral fluid therapy, bolus colloid infusions and/or blood products, together with sophisticated inten-

sive care management of the complex complications that often accompany severe shock. Of note, fluid overload is a significant contributor to morbidity and mortality in these circumstances, and balancing parenteral fluid therapy at a level just sufficient to maintain cardiovascular stability and critical organ perfusion during the phase of vascular leakage requires considerable skill and experience. One challenge when considering fluid replacement for dengue, especially DSS, is the current controversy regarding the safety and efficacy of fluid regimens using hydroxyethyl starch solutions, the only type of colloid which is affordable and available in many dengue-endemic countries, in the context that there are increasing concerns about the safety of this colloid in patients with severe sepsis (Maitland et al., 2011; Myburgh and Mythen, 2013; Huynh et al., 2013).

Regarding prevention, there is still no licensed dengue vaccine available although considerable research efforts have been dedicated to this field since the first attempts more than 65 years ago (World Health Organization, 2012d; Coller et al., 2011). Difficulties that have hampered the development of vaccines for dengue include the fact there are multiple serotypes, the lack of reliable animal models and the potential to induce an immune response which could lead to severe dengue after vaccination (Coller et al., 2011). However, advances in basic science have led to significant progress in the last decade with at least 5 vaccines in the clinical development stage and several others in preclinical stages (Coller et al., 2011). Recent publications regarding the most advanced dengue vaccine, the live-attenuated tetravalent dengue vaccine developed by Sanofi Pasteur, indicate that the vaccine has a good safety profile and significant efficacy. Two large-scale randomized, observer-blind, controlled multi-center phase III trials have reported relative risk reductions for symptomatic infection between months 13 and 25 after the first vaccine injection of 56% in more than 10,000 volunteers in Asia and 61% in more than 20,000 volunteers in Latin America (Capeding et al., 2014; Villar et al., 2014), although the findings from the initial phase IIb trial were not very convincing (Sabchareon et al., 2012). Even though these results are promising, there are still open questions regarding the longer term effects of the vaccine on the risk of severe dengue, the heterogeneity in serotype-specific efficacy, and the lack of efficacy in cases naïve to dengue that need to be addressed before introducing this vaccine into national vaccination programs (Wilder-Smith, 2014).

1.2 Factors associated with severe dengue

The outcome of dengue in an individual appears to be determined by a complex interplay of viral and human factors. This section describes a number of factors which have been presented in the literature (see reviews by Pawitan (2011); Yacoub et al. (2013); Whitehorn and Simmons (2011); Srikiatkhachorn and Green (2010); Huy et al. (2013b)), and are considered to play an important role.

1.2.1 Viral determinants

Viral factors which may be associated with dengue severity include viral load and viral serotype (Vaughn et al., 2000), with DENV-2 and DENV-3 possibly associated with severe dengue (Endy, 2002; Nisalak et al., 2003), although severe disease has been reported with all four serotypes. There is also some evidence that virus genotypes within the same serotype might have different virulence (Rico-Hesse et al., 1997).

1.2.2 Host determinants

Epidemiological factors

- Gender: results from epidemiological research in Asian countries suggests that amongst hospitalized dengue patients, females are more likely to develop DSS or die (Huy et al., 2013b). Possible explanations include a discrepancy in the healthcare-seeking behavior between genders in Asian countries or a greater susceptibility to capillary leakage in females (Anders et al., 2011).

- Age: patients at the extremes of age (young children and the elderly) have a higher risk of DSS and mortality (Anders et al., 2011). The greater risk of severe dengue in children may be explained by the higher intrinsic permeability of vascular endothelium in the young (Gamble et al., 2000).

- Comorbidity: patients with pre-existing conditions who may have underlying microvascular damage such as diabetes mellitus, hypertension and renal failure seem to have higher risk of severe dengue (Pang et al., 2012; Figueiredo et al., 2010; Kuo et al., 2008).

Clinical signs and symptoms

- Clinical signs and symptoms that are associated with severe dengue include pleural effusion, ascites, abdominal tenderness, hepatomegaly, lethargy, gastrointestinal bleeding, vomiting (Branco et al., 2014; Wichmann et al., 2004; Gupta et al., 2011). These associations could be explained by the relationship of these signs with plasma leakage, haemostatic dysfunction and damage in systemic organs.

- Laboratory findings which are often observed in dengue, and which tend to be increasingly abnormal with increasing disease severity, include thrombocytopenia (decreased level of platelet count (PLT)), haemoconcentration (increased level of haematocrit (HCT)), coagulation derangements, and increased hepatic transaminases (Chuansumrit et al., 2010; Tantracheewathorn and Tantracheewathorn, 2007; Wichmann et al., 2004; Tee et al., 2009; Almas et al., 2010). Disturbance of these laboratory tests in severe dengue have been attributed to excessive plasma leakage in combination with the overproduction of cytokines. Recent evidence also suggests that platelet activation could contribute to the elevated vascular permeability in dengue infection through the releasing of IL-1β (Hottz et al., 2013).

Host humoral immune response

- The effect of the host's humoral immune response on severity of disease is demonstrated by the positive relationship between secondary infection with a heterotypic serotype and severe disease (Sangkawibha et al., 1984; Chau et al., 2008; Halstead et al., 1970; Kliks et al., 1988). The classical explanation for this relationship is the ADE, a theory which suggests that non-neutralising antibodies, elicited in response to the previously encountered serotype, contribute to an increase in virus-infected cells during the current infection (Halstead and O'Rourke, 1977).

Host cell-mediated immune response

- The magnitude of the T-cell response influences severity of disease due to the excessive production of pro-inflammatory cytokines during T-cell activation (Duangchinda et al., 2010; Mongkolsapaya et al., 2003). In addition, high concentrations of several

cytokines including TNF-α, IL-1, IL-6, IL-10 and chemokines produced by dengue-infected cells might also be associated with severe dengue (Butthep et al., 2012; Appanna et al., 2012).

Host genetics

- Patients with African ancestry may have a lower risk for severe dengue (Sierra et al., 2007). Furthermore, several genes related to disease susceptability have been identified including MICB, PLCE1 (Khor et al., 2011), and other genes encoding a number of blood antigens and immune effector cells (Sakuntabhai et al., 2005; Vejbaesya et al., 2009).

1.3 Prognostic models in dengue

1.3.1 Introduction to prognostic models

Prediction/prognosis in clinical medicine

In daily clinical practice, one of the main tasks of clinicians is to investigate and develop three main areas of knowledge related to the patient's illness: diagnosis (whether a particular illness is present), aetiology (what is the cause of the present illness) and prognosis (what is the likely future course of the present illness) (Miettinen, 2011). Based on this, clinicians can suggest a suitable intervention or management plan for the patient under investigation. Therefore, making prognostic assessments is a natural and familiar task that clinicians have to do every day.

As prediction is a fundamental component of medical decision-making, it has an important role in clinical practice. However, it is also a generally difficult task because unlike diagnosis or etiognosis, prognosis requires extrapolation to the future based on the present knowledge about a patient's disease status and characteristics. As the future is uncertain, predictions need to rely on many assumptions and making a good prediction is an extremely difficult task.

In general, the prediction making process may come with pitfalls including its subjectivity, proneness to errors, and inconsistencies (Meehl, 1954; Dawes et al., 1989; Grove

et al., 2000; Sox et al., 2013). Common pitfalls when estimating probability using only personal experience are:

- Focusing on the presence or absence of predictors but ignoring the prior probability of the outcome.
- Basing predictions on the presence of predictors with low predictive ability or a set of predictors which do not independently affect outcome.
- Personal experience, especially for clinicians in their early careers, is typically a small and unrepresentative sample of the overall population.
- Reliance on a false belief in the relationship between candidate predictors and outcome, as it is difficult to differentiate invalid relationships from valid ones theoretically.
- A tendency to be over-confident and overstate subjective certainty.

As a result, deriving precise and unbiased predictions from personal experience alone is beyond the cognitive ability of almost everyone (Sox et al., 2013). Therefore, additional knowledge, especially that deriving from an objective source like empirical research, is needed. Such knowledge includes published reports describing the incidence of outcomes amongst patients having a common set of clinical features and clinical prediction models. Compared to personal experience, they provide more objective and more comprehensive information, especially for rare outcomes. However, their utility and validity can vary widely depending on the quality of the clinical studies and the similarities between the research population and the population of interest. Amongst these objective sources of knowledge, clinical prediction models are recognized as powerful tools to derive prediction. Besides their ability to process and produce complex information that may go beyond human mental ability, clinical prediction models provide consistent estimates and have been shown to outperform personal judgement in many situations (Dawes et al., 1989; Grove et al., 2000). Caveats in applying clinical prediction model are related to their potential to over-fit the data when they are not developed or tested properly and their potential to miss predictors which are relevant but difficult to evaluate in clinical

research. For more information on what prognostic models are and how they are derived I refer to the following sections and Chapter 2.

Based on these rationales, a better model for estimating the risk of the occurrence of a certain outcome in clinical practice requires both published evidence and personal experience. In addition, further adjustments to risk estimation may be required due to differences between the population of interest and the reference population, or due to the presence of extraordinary clinical features in the patient that the physician has never seen and which have not been reported anywhere (Sox et al., 2013).

What are prognostic models?

A prognostic model is a formal combination rule of multiple predictors from which a subject's absolute risk of the occurrence of a disease event of interest can be calculated (Steyerberg et al., 2013). Prognostic modeling is an important part of prognostic research which aims to understand and improve future outcomes in subjects with a given health state (Hemingway et al., 2013). Developing a prognostic model is the third step of a 4-step paradigm for prognostic research which includes (1) investigating the variation of outcomes of a health condition in the context of current care (fundamental prognostic research); (2) identifying risk factors which are related to outcomes (prognostic factor research); (3) developing, validating and assessing the impact of prediction models that predict an individual's risk of a future event (prognostic model research); (4) using prognostic information to help individualize treatment decisions for a subject or group of subjects that share similar characteristics (stratified medicine research) (Hemingway et al., 2013).

A typical prediction model has three main ingredients: an outcome, candidate predictors, and a statistical model. An outcome can be a hard endpoint such as death or the presence of clinical complication, or a surrogate endpoint such as a biomarker of severity, or a composite endpoint. Candidate predictors typically include demographic variables such as age and gender as well as clinical symptoms or biomarkers which are relevant to outcome prediction based on clinical knowledge. As prognosis is aimed at the future, predictors have to be collected at a starting point or baseline which is before the outcome occurs and the length of the lag period between the starting point and the outcome oc-

currence affects largely the usefulness of the derived prediction model. The relationship between outcome and predictors is modeled using a statistical model and the choice of this model depends on the type of outcome. As clinical outcomes are usually binary or survival data, the most common statistical models of choice are logistic regression and the Cox proportional hazards model, but many other statistical models have been suggested in the literature (Steyerberg, 2010; Hastie et al., 2009). As a general rule, these three main components have to be pre-specified before fitting any model in order to avoid over-fitting and therefore to preserve the validity of the derived model.

Prediction models are different from decision rules. The inputs to a prediction model are values of prognostic factors at a pre-defined time point and the output is an estimated risk of a specific outcome. Even though one may categorize estimated risks and assign suggested actions to each risk category, a typical prediction model only provides a prediction of a risk as it is intended to assist clinicians without suggesting to them what to do (Reilly and Evans, 2006). In contrast, a decision rule is designed to directly affect clinical decisions by physicians. As accurate predictions do not always improve clinical decisions, a promising prediction model has to demonstrate its positive impact on physicians' decisions and patients' outcomes through different levels of impact analysis in order to be successfully translated into a useful clinical decision rule (Reilly and Evans, 2006).

As described in the previous section, the major advantage of prediction models in clinical practice is their ability to provide objective, reproducible and reliable estimation of outcome occurrence. Due to their transparency, prediction models can also enhance communication between physicians and patients (Steyerberg et al., 2013). Moreover, in clinical research prediction models can be used in the design stage to target a population of interest, and in the analysis stage to perform stratified analysis according to predicted risk groups or to improve the power of statistical analyses (Steyerberg et al., 2013).

Baseline versus dynamic prediction models

Traditionally, a prediction models relies on data collected at a single time point (at the time of presentation, admission, diagnosis or initiation of an intervention) to predict outcomes in the future. Even though many of these traditional prediction models are useful in clinical practice, they have several shortcomings:

- Initial predictions tend to become less relevant as the disease progresses (Rué et al., 2001). A possible explanation is that a prediction model based on baseline information only cannot capture changes in the patient's clinical profile according to their response to treatment or natural physiologic variation reflecting the course of the disease, which may be strong predictors of the outcome (Rué et al., 2001). Furthermore, baseline models might also miss complications, which may strongly affect outcome, while being present at baseline but require time to become clinically apparent (Lemeshow et al., 1988). As a result, baseline models might not be transferable to later time points and therefore, they may not be used for individual management decisions at those time points (Lemeshow et al., 1994; Wagner et al., 1994) .

- Longitudinal information during the patient's disease course is nowadays frequently collected in clinical practice, especially with the introduction of electronic health record. Baseline models are inefficient in the sense that they ignore all this post-baseline information.

For these reasons, dynamic prediction models, which predict the future course of the disease at follow-up time points based on the accruing longitudinal information, are required to allow updating a patients' prognosis over time (Van Houwelingen and Putter, 2012). By using all available data, such models may provide much more accurate predictions compared to baseline models in many settings (Lemeshow et al., 1988; Christensen et al., 1993; Hughes et al., 1992; Rué et al., 2001; Karp et al., 2004). Dynamic prediction may also be appealing for clinicians as it mimics the iteration of obtaining information and updating prognosis based on the new information, a task that physicians routinely do every day in clinical practice.

1.3.2 Prediction models in dengue

As described in the previous sections, there is still no specific treatment for dengue and case management relies mainly on supportive treatment. Therefore, improving outcomes in patients with dengue depends in part on effective triage to identify patients who are likely to progress to more severe disease at an early stage, and reliable prediction models could facilitate this. As a first step, exploring the current status of prognostic research in

the dengue field should provide useful information to justify and tailor the specific needs for developing prediction models in dengue.

Literature Review

A PubMed search for prediction models in dengue on 10 September 2014 yielded 263 articles amongst which 17 original articles described actual prediction models (Table 1.1). These studies were published from 2005 to 2013 (Ibrahim et al., 2005; Iskandar et al., 2008; Lee et al., 2008; Tanner et al., 2008; Lee et al., 2009; Gomes et al., 2010; Ibrahim et al., 2010; Potts et al., 2010a,b; Thein et al., 2011; Brasier et al., 2012b,a; Faisal et al., 2012; Dewi and Nurfitri, 2012; Ju and Brasier, 2013; Pongpan et al., 2013; Huy et al., 2013b).

Most of these studies utilized prospective data (13/17) whereas the remaining 4 either used retrospective data (3 studies) or did not clearly specify the type of design (1 study). Most of them were conducted in South-East Asia (13/17) while 4 were from Latin America. The studies included only children (6/17), only adults (4/17), children and adults (5/17), or the age of participants was not specified (2/17). Patients were enrolled from hospitals (12/15) or community clinics (3/15), and most studies only included subjects with laboratory-confirmed dengue (15/17). The majority of studies focused on predicting severe outcomes in patients with dengue infection (14/17) whereas 3/17 studies aimed to predict severe outcome in patients with DSS. Amongst the 3 studies in patients with DSS, 2 studies aimed to demonstrate the applicability of established general severity scores from intensive care medicine (PELOD, PRISM) to patients with DSS, while only 1 study aimed to develop a novel prediction model for severe outcome in patients with DSS.

Regarding clinical outcomes, 2 studies defined multiple endpoints whereas all others only used a single endpoint. Amongst patients with dengue infection, severe outcomes were defined as DHF, DSS, or DHF plus other severity criteria concerning pleural effusion or specific laboratory values (PLT, HCT, aspartate aminotransferase (AST)). Two studies also used PLT counts on day 5 to 7 of illness, and the day of defervescence, as surrogate markers for severity in these patients. Amongst patients with DSS, severe outcomes included mortality and recurrent shock.

Most studies used a large number of candidate predictors, ranging from general demo-

graphical factors to clinical symptoms, signs and common laboratory tests. More advanced and complicated factors that were occasionally included were proteomic or gene expressions profiles and bioelectrical impedance measurements. In two studies, well-known severity scores to predict mortality in children in critical care units were directly assessed. In most studies, values for the candidate predictors were collected at the time of presentation. In 3 studies, repeated measurements of predictors were also included; however, data from these repeated measures were simply pooled without taking into account the dependence of measurements from the same subject. To estimate the effect of candidate predictors on outcome, these studies had sample sizes varying from less than 100 patients (6/17) to 100-1000 patients (5/17) and 1000-2000 patients (6/17). The effective sample size, or the number of subjects experiencing the events of interest, also fluctuated widely from 5 to 228 patients. In particular, the effective sample sizes for the studies aiming to predict development of DSS amongst patients with dengue were quite low (from 37 to 90 observed DSS cases).

Logistic regression modeling was the most common statistical model used to derived these prediction models (5/14 studies). Other statistical methods were classification and regression trees (CART), artificial neural network (ANN), support vector machine (SVM) and multivariate adaptive regression splines (MARS). Three studies also compared the performance of different statistical modeling approaches as part of their model development.

Three of the 17 studies aimed to externally validate previously derived prediction models. Amongst the others, only one prediction model was validated externally using an independent dataset, 6 were validated internally using cross-validation and sample-splitting methods, and 7 were evaluated in the original dataset without any adjustment. Methods for dealing with missing data was described in only 5 studies, while only 1 study reported an assessment of the plausibility of modeling assumptions.

Table 1.1. Main characteristics of the 17 articles included in this review.

Article	Study design	Hospital/ Community	Study population	Age	Sample size	Outcome	Candidate predictors	Statistical model	Validation
Ibrahim et al. (2005)	Prospective	Hospital	DF	Adults	252	Day of defervescence	Clinical symptoms	ANN	NS
Iskandar et al. (2008)	NS	Hospital	DSS	Children	41	Mortality	PELOD score, PRISM III score	No	External
Tanner et al. (2008)	Prospective	Hospital	DF	Adults, children	161	PLT on day 5-7 $< 50{,}000$	PLT, Ct[a] IgG (+)	CART	Internal
Lee et al. (2008)	Retrospective	Hospital	DF	Adults	1937	DHF	Demographic, clinical symptoms, laboratory tests	Logistic regression	No
Lee et al. (2009)	Retrospective	Hospital	DF	Adults	1937	DHF	Demographic, clinical symptoms, laboratory tests	CART	Internal
Gomes et al. (2010)	Prospective	Hospital	DF	Adults, children	28	DHF	Expression pattern of 12 genes	SVM	Internal
Ibrahim et al. (2010)	Prospective	NS	DF	NS	207	Low risk vs high risk	Day of fever, gender, reactance value	ANN	No
Potts et al. (2010a)	Prospective	Hospital	Febrile $\leq$ 72 hours	0.5-15 years	1230	(1) DSS, (2) DSS or dengue with significant pleural effusion	Demographic, laboratory tests	CART	Internal

Potts et al. (2010b)	Prospective	Hospital	Febrile $\leq$ 72 hours	0.5-15 years	1227	(1) DHF vs DF, (2) DHF vs (DF + OFI), (3)(DHF + DF) vs OFI, (4) evere dengue (DSS + PEI > 15 + required too much fluid + required IV fluid) vs (non-severe + OFI)	Laboratory tests	Logistic regression	External
Thein et al. (2011)	Prospective	Hospital	DF	Adults	1017	DHF	Previous model	Logistic regression	External
Dewi and Nurfitri (2012)	Prospective	Hospital	DSS	Children	81	Mortality	PELOD score		External
Brasier et al. (2012b)	Prospective	Community	DF	Adults, children	51	DHF	Clinical symptoms, laboratory tests	Bayesian feature reduction, logistic regression, CART, RF, GAM	No
Brasier et al. (2012a)	Prospective	Community	DF	Adults, children	55	DHF	Clinical symptoms, cytokine profiles, proteomics	MARS	No
Faisal et al. (2012)	Prospective	NS	DF	NS	210	Low risk vs high risk[b]	Physiology profiles, clinical symptoms, bioelectrical impedance analysis	ANN	No
Pongpan et al. (2013)	Retrospective	Hospital	DF	1-15 years	777	DF or DHF or DSS	Demographic, mode of presentation, haemodynamic profiles, haematological profiles, biochemical profiles	Ordinal logistic regression	No

Ju and Brasier (2013)	Prospective	Community	DF	Adults, children	51	DHF	Biochemistry profiles, cytokine profiles	SAM, TreeNet, BMA, SSVS, GPS, MARS, RF	Internal
Huy et al. (2013b)	Prospective	Hospital	DSS	0.5-15 years	444	Recurrent shock	Demographic, clinical symptoms, laboratory tests	Logistic regression	Internal

All studies (except for Gomes et al. (2010) and Pongpan et al. (2013)) only included patients with laboratory-confirmed dengue.

[a] *Ct = the cross-over value of the real-time RT-PCR for dengue viral RNA.*

[b] *Low risk = DHF with < 2 severe criteria, High risk = DHF with ≥ 2 severe criteria (severe criteria = PLT $\leq$ 30000; HCT $\geq$ 20%; AST or ALT increased by 5-fold).*

Abbreviations: PEI = pleural effusion index, IV = intravenous, OFI = other febrile illness, DF = dengue fever, DHF = dengue haemorrhagic fever, DSS = dengue shock syndrome, PLT = platelet count, HCT = haematocrit, AST = aspartate aminotransferase, ALT = alanine aminotransferase, ANN = artificial neural network, CART = classification and regression trees, SVM = support vector machine, RF = random forest, MARS = multivariate adaptive regression splines, SAM = significance analysis of microarray, TreeNet = learning ensemble, BMA = Bayesian model averaging, SSVS = stochastic search variable selection, GPS = generalized path seeker, PLT = platelet count, NS = not specified.

Summary

The above review of the prediction modeling in dengue shows that although the topic is increasingly attracting attention, most of developed models have certain limitations. Several important areas to be addressed are as follows:

- Modeling strategy: the current models were developed using a variety of different modeling strategies, but many of them failed to take into account the importance of missing values, modeling assumptions or variable selection. Furthermore, the risk of over-fitting was not addressed properly as many studies still assessed performance of the derived model using the original data.

- Study population: most studies focused on predicting severe outcomes in patients with dengue, and used relatively small datasets. There were only 3 studies that looked at outcomes in patients presenting with established DSS.

- Clinical outcomes: there are currently no standardized clinical outcomes in dengue; therefore, endpoints vary considerably between studies and formal comparisons are difficult. Amongst patients with dengue, the most common outcome was development of DHF; however, over a number of years increasing concerns have been raised regarding the complexity and usefulness of the DF/DHF classification system, in particular the requirement for four specific criteria to support a diagnosis of DHF such that some patients with clinically severe disease are categorized inappropriately. Amongst patients with DSS, identified outcomes were mortality and recurrent shock; however, in the absence of a comprehensive description of patients with DSS or the frequency of relevant outcomes, it is hard to decide which would be most relevant for a reliable prediction model.

- Utilization of repeated measurement: most of the prediction models developed to date used only baseline information for the candidate predictors. Three studies included repeated measurement but failed to take into account the dependency between measurements on the same subject. As repeated measures, e.g. changes in HCT or PLT, might carry important information on a subject's prognosis, there is a need to develop dynamic prognostic models which exploit this and, where necessary,

to develop the required statistical methodology to achieve this goal.

1.4 Aims of the thesis

In the context of the limitations of current prognosis research in dengue and with the availability of an extensive resource of more than 4000 well-characterized patients entered into various dengue studies over the last 10 years at OUCRU, in this PhD project I have set out to develop robust prediction tools for use in clinical practice, and to improve methodology in building dynamic prediction models which incorporate longitudinal information in the acute disease setting. Specifically, the aims of this work can be broadly separated into two main areas - clinical, and statistical.

The clinical aims of my PhD thesis are:

1. To describe the clinical and laboratory features of DSS in children,
2. To develop prognostic models for severe outcome amongst children with DSS, using baseline information obtained at the time of development of shock,
3. To develop prognostic models for progression to DSS in children hospitalized with dengue using both baseline and longitudinal information.

The statistical aim is to adapt and compare available methods for developing and assessing dynamic prognostic models using the datasets utilized to achieve the clinical aims.

Based on these aims, the thesis is structured as follows:

- Chapter 2 describes the clinical datasets, study procedures and common analytic methods used in this thesis (descriptive analysis, treatment of missing values, building blocks of prediction model development using baseline information only),
- Chapter 3 describes the characteristics at presentation with shock, and the clinical evolution during hospitalization, of over 1700 children with DSS,
- Chapter 4 assesses risk factors for severe outcome amongst children with DSS and presents a prognostic model for severe outcome using baseline information only,

- Chapter 5 assesses risk factors for DSS development amongst children hospitalized with dengue and presents a prognostic model for DSS using baseline information only,
- Chapter 6 describes the current knowledge about dynamic prediction models, implications of differences between acute and chronic diseases on dynamic prediction modelling, and compares methodology to develop and evaluate such models, using dengue as a case study,
- Chapter 7 summarizes the conclusions from this thesis and suggests a number of avenues for future research in this field.

1.5 Appendix

Search term for the literature review of prognostic models in dengue ((("dengue hemorrhagic fever"[MeSH Terms] OR "dengue shock syndrome"[Text Word]) OR "dengue shock"[All Fields]) OR (severe[All Fields] AND ("dengue"[MeSH Terms] OR "dengue"[All Fields]))) AND (("Decision Support Techniques"[Mesh] OR "Bayes Theorem"[Mesh] OR "Prognosis"[Mesh] OR "Forecasting"[Mesh]) OR "Risk Factors"[Mesh]).

Chapter 2

Materials and common analytical methods

Summary

This chapter describes the datasets used in this PhD project, together with detailed information regarding data collection and laboratory diagnostics. In addition, this chapter specifies common analytical methods used in the subsequent chapters, including descriptive analyses, treatment of missing values and the chosen strategy to develop prognostic models using baseline information.

2.1 Materials

In this PhD project I utilized data from two large paediatric cohorts (with study codes DF and MD) to describe clinical features of the corresponding populations of children with dengue, to develop prognostic models for severe outcomes, and to illustrate the development of dynamic prognostic models for acute diseases.

2.1.1 Study populations

DF cohort

This prospective cohort was enrolled at the paediatric intensive care unit (PICU) of the Hospital for Tropical Diseases (HTD) in Ho Chi Minh City, Vietnam, from 1999 and 2009. The participants comprised children below 15 years of age admitted to the PICU with a clinical diagnosis of DSS. To limit the effect of treatment before admission on outcome, patients transferred from other facilities for tertiary care (after initial shock resuscitation) were not included. However patients transferred from other wards at HTD after development of DSS, some of whom had received some maintenance fluid therapy during the febrile phase, were eligible for enrolment.

From 1999 to 2009, 1810 children with DSS were enrolled in this cohort. For each patient, information on baseline characteristics at enrolment (collected within 2 hours of presentation with DSS), detailed information on treatment during hospitalization, and clinical outcomes at discharge were collected. Daily PLT were measured and HCT assessments were repeated every 2 to 12 hours with the exact schedule depending on the clinical status of the patient.

During the first 6 years of the study (1999 to 2004), 503 patients in this cohort were also recruited into a nested randomised controlled trial (RCT) which compared different fluid types for initial resuscitation in children with DSS (Wills et al., 2005). All patients enrolled during 1999-2002 were recruited into this trial; however, a number of patients were not included in the RCT during 2003-2004 as the trial was temporary stopped due to safety reasons.

MD cohort

This prospective cohort was recruited at HTD between 2001 and 2009. Participants comprised children aged between 5 and 15 years admitted to the dengue ward with a clinical suspicion of dengue. In total, 3040 patients were enrolled into this study. The available data included baseline information at enrolment, treatment given, and clinical outcomes at discharge, as well as daily HCT and PLT values.

2.1.2 Study procedures and data collection

In both cohorts, trained study doctors obtained written informed consent from a parent or guardian (after giving verbal and written explanations), and then enrolled the children into the study. At enrolment, baseline information including demographic characteristics, clinical history, and examination findings were collected using a structured case report form (CRF). During hospitalization, all patients were followed daily and information on therapeutic interventions and supportive care, especially on fluid usage and major clinical events, were recorded by study doctors. At discharge, study records were reviewed and each patient was assigned a final diagnosis and clinical disease category based on WHO and local clinical guidelines (World Health Organization, 1997). All patients were asked to return for follow-up assessments at one month, and again at two months if there were any ongoing concerns. In the DF cohort, the original study CRF contained only limited information on fluid interventions. Therefore I designed an additional short CRF to collect more detailed information on fluid usage, and worked with two study nurses to extract relevant data from the hospital files and other source documents.

For all patients, 5 ml venous blood samples were obtained on the day of enrolment, the day of discharge or defervescence, and at the follow up visits. These blood specimens were used to perform dengue serology and reverse transcriptase polymerase chain reaction (RT-PCR) tests. In addition, 1 ml finger prick blood samples were obtained daily (in MD) or at varying intervals from 2-12 hourly (in DF) as part of standard care for dengue patients at HTD. These blood samples were used to measure HCT levels and PLT counts of the patient at that time point.

During the study period, patients were managed following the treatment guidelines

of HTD and supervised by a group of senior clinicians which remained stable during the whole study period. In the DF cohort, initial treatment in the first two hours after presentation with compensated shock (for patients not enrolled in the RCT) was 25 ml/kg of Ringer's Lactate fluid, or the same volume of a colloid solution (6% dextran or starch) for patients presenting with decompensated/hypotensive shock. Subsequently, a standardized schedule of Ringer's Lactate was used, involving staged reductions at specific time intervals, aiming for maintenance fluid therapy after eight hours. Patients in the RCT were randomized to receive one of three possible intravenous fluid solutions at a rate of 25 ml/kg over 2 hours for initial volume resuscitation, but subsequent management, study procedures and data collection were otherwise similar to other patients in the cohort. Patients whose cardiovascular status failed to stabilize within the first two hours or who deteriorated during the mandatory 36-48 hour period of close observation received 10-15 ml/kg infusions of rescue colloid plus inotropes, blood products or other therapies at the discretion of the treating clinician.

Both cohorts were approved by the HTD ethical committee and the Oxford Tropical Research Ethics Committee.

2.1.3 Laboratory diagnostics

In this project, two sets of laboratory diagnostics were performed: (1) to confirm the dengue diagnosis, and (2) to differentiate primary and secondary dengue infections. Only patients with laboratory-confirmed dengue were included in the main analyses.

Determination of dengue diagnosis

A laboratory-confirmed case was defined by detection of DENV RNA in plasma (PCR), or by seroconversion on the capture ELISA (ELISA). In the DF study, patients with elevated dengue-specific IgM at onset of shock were also considered as confirmed dengue cases provided the overall clinical picture was consistent with DSS. Cases were defined as dengue-negative if the enrolment RT-PCR and paired serology specimens were all negative.

ELISA Dengue IgM and IgG capture ELISAs were performed on paired enrolment and early convalescent specimens. During the study period the diagnostic laboratory used a number of different serological tests, following the manufacturer's instructions for commercial kits (Dengue Duo IgM and IgG Capture ELISA, PanBio, Australia), or established standard operating procedures for in-house methods (Cardosa et al., 2002).

Cross reactivity with flaviviruses that co-circulate in the region may influence the results of IgG serology in particular in Vietnam, Japanese encephalitis (JE) virus is known to circulate and to cause sporadic cases of meningo-encephalitis. However, although an inactivated JE vaccine was introduced in 1997 for use in high-risk areas, southern Vietnam is not considered to be high-risk and JE vaccination is not part of the local Expanded Program of Immunization. Vaccination became available privately in Ho Chi Minh City from around 2005-2006 onwards but uptake remains sporadic. Although the specific sero-diagnostic tests performed did not assess cross-reactivity for JE virus, due to the low vaccine coverage locally it is unlikely that JE vaccination affected the identification of secondary dengue cases in this study. Symptomatic disease caused by JE virus is primarily neurological and unlikely to mimic dengue infection. None of the patients in either cohort had a past history of serious neurological disease, but it is possible that recent asymptomatic or pauci-symptomatic JE virus infection influenced the serological responses we documented, although the number of cases affected is likely to be small.

PCR RT-PCR was performed on the enrolment specimen using established methodology (Lanciotti et al., 1992; Shu et al., 2003a).

Determination of primary/secondary dengue infection

As described in Chapter 1, serological definitions for primary versus secondary infections commonly rely on the ratio of IgM/IgG, but may give varying results depending on the test used and the day of illness when the specimen is obtained. Given that a number of different sero-diagnostic tests were employed during the 10-year study-period serologic classification of dengue infection was determined using the following simple definitions:

- Primary infection: if the patient had two negative dengue-specific IgG results provided that the second sample was obtained during the second week of illness.

- Secondary infection: if there was at least one positive dengue-specific IgG on or before day 7 of illness.
- Possible primary infection: if the patient only had single negative dengue-specific IgG, either in the first or the second week of illness.
- Unclassifiable: all other patients, the primary reason for being 'unclassifiable' was the availability of a single specimen or late convalescent specimens only.

Of note, as the level of dengue-specific IgG increases over time during dengue infection, there would be a high chance that a patient with primary dengue infection has positive IgG result in the late course of illness (i.e. day 6, day 7 of illness). Therefore, this simple classification is expected to provide a relative accurate detection of primary dengue infection, but a less precise detection of secondary cases.

2.1.4 Data cleaning and checking

A team of experienced physicians and I were responsible for checking and cleaning the study databases. This procedure involved two steps: checking individual CRFs and checking the aggregated databases. In the first step, I and other physicians examined all written CRFs to check the consistency and accuracy between written CRFs and the electronic databases. In the second step, I checked the whole database to find outliers or implausible values and then traced them back to the written CRF for cross-checking if necessary. All corrections were documented in both the written CRFs and the electronic databases.

2.2 Common analytical methods

2.2.1 Descriptive analysis

The distribution of variables was described by numbers and/or graphs. Numerical summaries were median and the corresponding interquartile range (IQR) for continuous variables, or frequency and percentage for categorical variables. Graphical displays included histogram for continuous variables, and bar plots for categorical variables. The evolution of longitudinal data over time was visualized using plots of patient profiles over time together with a scatter-plot smoother based on local regression (Diggle et al., 2002).

2.2.2 Treatment of missing values

The extend of missing data was assessed for all pre-defined candidate predictors. The amount of missing values were summarized in terms of the frequency (%) of missing values per variable, the range of the total number of missing values per individual, and the frequency and fraction of incomplete cases (individuals with at least one missing value). To assess the plausibility of the missing completely at random (MCAR) assumption, I also investigated the relationship between indicators of missingness in a certain variable and the observed values of other covariates using multivariable logistic regression (Steyerberg, 2010).

To deal with missing values, common approaches are complete-case analysis, single imputation and multiple imputation. The decision regarding the method of choice depends largely on the suspected missingness mechanism and the amount of missing values in a specific situation (Steyerberg, 2010). A simple rule of thumb proposed by Harrell suggests using complete-case analysis or single imputation when the fraction of incomplete cases is less than 5%, single imputation or multiple imputation when the fraction of incomplete cases ranges from 5% to 15%, and multiple imputation when the quantity is larger than 15% (Harrell, 2001).

In the DF cohort, I decided to use a single imputation as the fraction of incomplete cases was low (4%, details are presented in Chapter 4). Specifically, missing values were imputed with the median of non-missing values for continuous variables, or the most frequent category for categorical variables. However, univariate analyses were still based on complete-case analyses.

In the MD cohort, the fraction of incomplete cases was higher (7%) and there was also evidence that the missingness of certain variables depended on observed values of other covariates, which implies that the MCAR assumption might be violated (details are presented in Chapter 5). Therefore, multiple imputation was chosen to deal with missing data.

Specifically, I used multivariate imputation by chained equations (MICE) as implemented in the R package mice version 2.22 (Van Buuren and Groothuis-Oudshoorn, 2011), to generate multiple imputed data sets based on a set of imputation models (each variable

with missing values has one imputation model). MICE is an iterative procedure and in the first step, all missing values were imputed by values randomly chosen from observed values. For the first variable with missing values, the parameters of its imputation model were then estimated based on individuals with non-missing values for that variable and then the models' posterior predictive distribution was used to draw imputed values for missing data. In an iterative fashion, this procedure was repeated for all variables in a pre-defined order and many iterations were performed to create a single imputed dataset in order to stabilize the result. The quality of the whole imputation process was then assessed by examining variances within and between parallel imputation streams, and the distribution of imputed values (Van Buuren and Groothuis-Oudshoorn, 2011).

For the MD data, the chosen imputation models were predictive mean matching for continuous variables, logistic regression for binary variables and multinomial regression for categorical variables with more than 2 classes (Van Buuren and Groothuis-Oudshoorn, 2011). The advantage of predictive mean matching is that imputed values match actually observed data which might be more appropriate than regression imputation if the normality assumption is violated. As recommended, these models included outcomes and all candidate predictors (with only linear terms for continuous covariates and no interaction terms) (White et al., 2011). The visit order of the variables in each iteration cycle was according to their (increasing) number of missing values. In total twenty imputed datasets were created and 50 cycles per dataset were performed. Of notes, this exceeds the minimum required numbers of imputed datasets and repeated cycles according to current recommendation which are 7 (the percentage of incomplete cases) and 10-20, respectively (White et al., 2011).

These imputed datasets were used throughout the whole analyses except for univariate analyses which were based on complete-case analyses. Estimates and asymptotic covariance matrices (and associated Wald-type tests) were combined across multiple imputed datasets using Rubin's rule and likelihood ratio tests for multiple imputed datasets were calculated using the method of Meng and Rubin (Meng and Rubin, 1992; White et al., 2011). As recommended, I also performed complete-case analyses (in all steps) and examined differences between results (Sterne et al., 2009).

2.2.3 Strategy to develop prediction models using baseline information

In this project, all prediction models using baseline covariates were developed following current standard methodology and recommendations (Harrell, 2001; Steyerberg, 2010).

Clinical outcomes and candidate predictors

Depending on the specific context of each study, clinical outcomes and candidate predictors were pre-defined based on clinical knowledge and prior to any analysis. All values of candidate predictors were obtained at baseline. All cases in whom the outcome had already occurred at or prior to the baseline assessment were removed from the prediction model development.

Statistical models of choice

The primary outcomes in both cohorts were binary: severe DSS (yes/no) in the DF cohort, and DSS (yes/no) for MD. Hence, the logistic regression model was the statistical model of choice. In the MD data set, the time point of the occurrence of DSS was also recorded; therefore, in theory, time to DSS occurrence could also be used as an outcome and a survival model could be applied to develop a prognostic model. However, using time to event as an outcome requires a meaningful time origin. In this setting, two time origins could be used including the time point of enrolment and the time point of disease onset. As the former time point is quite arbitrary, it is less meaningful than the time point of disease onset. However, using the time of disease onset as the time origin posed two problems. First, subjects only came under observation at the time of enrolment and hence, the dataset contains possibly informative left-truncation. Second, several of the selected candidate predictors are in principle time-varying but for the majority of them their values were only recorded at a single time point, i.e. the time of enrolment. Thus, in order to keep the analysis simple and transparent, I decided to use the logistic regression model as the main statistical model for the analysis of MD data.

However, the validity of the logistic regression models is based on the assumptions of linearity and additivity of covariate effects while these assumptions are relaxed in several modern approaches. In addition, some of the more modern approaches are also less

prone to over-fitting. Therefore, I also applied modern statistical models including the lasso, generalized additive models (GAM), classification and regression trees (CART), and gradient boosting with trees as base learners (Hastie et al., 2009) and compared their performances to the main logistic models in order to detect any defects in the latter.

Model specification

For simplicity, the initial prediction model included all candidate predictors as linear and additive terms. These assumptions were subsequently assessed based on a pre-defined maximum amount of flexibility ("degrees of freedom") allowed for each continuous variables and pre-defined interaction tests. These pre-specifications depended on clinical judgement, expected associations from the literature, and the number of effective events in the data.

Model assumption assessments

For logistic regression models, the initial simple models were first assessed for the plausibility of common model assumptions.

Linearity assumption This assumption states that the effect of a candidate predictor on the outcome depends linearly on its value (the linearity is only applied to the appropriate scale of the model, e.g. linear on the log-odds ratio scale for a logistic model). In reality, this assumption is hardly ever completely true; however if the effect of a covariate on the outcome is approximately linear, using a linear model has the benefits of simplicity and transparency. However, the performance of a prediction model can be hampered when a truly non-linear relationship is forced to be linear. In this project, this assumption was assessed in two ways:

- Numerically by performing statistical test to compare goodness-of-fit between the initial model and a more flexible model which allows for non-linear effect. Natural cubic splines with pre-defined degrees of freedom and knot locations are often recommended for modeling non-linear effects (Harrell, 2001) and in this case, the linear and non-linear models can be compared using a likelihood ratio test.

- Graphically by assessing estimated non-linear effects of each continuous variable on the outcome from a flexible multivariable model which allows for non-linearity. The flexible multivariable model was chosen as a generalized additive model which included all continuous variables of interest modeled as natural cubic spline functions with automated selection of the required degree of smoothness, and the partial effect of each variable on outcome was extracted and visualized using term plots (Wood, 2006).

If pronounced non-linear terms were detected during this assessment, they were added to the model.

In contrast to classical linear regression models, GAM, CART and boosting with trees as base learners by default allow for non-linear covariate effects. However, in the case of CART, the linearity assumption is replaced by the often even less plausible assumption that the covariate effect can be described by a step function.

Additivity assumption The simple formulation of logistic regression also assumes additivity of covariates effects. This assumption is violated when there are (synergistic or antagonistic) interactions between covariates, i.e. if the effect of one covariate on the outcome depends on the levels of other covariate. Commonly seen interactions in clinical studies are between severity/place/time/age with other candidate predictors (Steyerberg, 2010). In my context, relevant potential interactions are between the day of illness at enrolment/gender and other covariates, and these interactions were assessed by overall interaction tests, i.e. likelihood ratio tests comparing the initial model and the extended model which also included pre-defined interaction terms. If this overall test was significant, further investigation was performed to identify the specific interaction. If pronounced interaction terms were detected during this assessment, they were added to the model.

Of note, the additivity assumption is relaxed in CART models and boosting with trees as base learners which automatically include interaction terms.

Model estimation

Parameters of logistic regression models were estimated using standard maximum likelihood estimation. Estimation of the penalty parameter for the lasso was based on standardized covariates and leave-one-out cross validation with the likelihood as the optimization criterion as implemented in the R package glmnet version 1.9.8 (Friedman et al., 2010). The CART model built and pruned back a classification tree using default parameter settings of the R package rpart version 4.1.8 (Therneau et al., 2014). The GAM model was built based on default settings of the R package mgcv version 1.8.3 (Wood, 2011). The implementation automatically estimates the degrees of freedom of smooth terms based on generalized cross-validation. To fit a "pure" additive model, the interaction terms were not included in the model formula. Finally, a generalized boosted regression model with a Bernoulli distribution for the outcome was fitted using classification trees as base learners as implemented in the R package gbm version 2.1.6 (Greg Ridgeway with contributions from others, 2014). Each tree has a depth of at most 2 which allows for 2-way interactions. The number of 3000 iterations and the learning rate of 0.001 were chosen as recommended by the gbm package author.

Model reduction

As some candidate predictors may have negligible effects on the outcome and the full model is generally complex, it is necessary to simplify the model before applying it to clinical practice (Steyerberg, 2010). In this project, I used several different variable selection techniques including stepwise selection using the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) as selection criteria and best subset selection which searches through all possible models to find the best one regarding AIC or BIC criteria. Of note, the lasso by default includes variable selection by shrinking coefficients of unimportant variables to zero.

Model performance

Performance criteria The performance of developed models was assessed in terms of overall performance, discrimination and calibration. These criteria are described in detail

in (Steyerberg, 2010) and only briefly discussed here. The overall performance of prediction models was quantified with the Brier score which is the average squared difference between patients' observed outcomes (0 for patients without the outcome, 1 for patients with the outcome) and their predicted risks. This quantity can range from 0 for a perfect model to a maximum value depending on the incidence of outcome for a non-informative model (for example 0.25 with a 0.5 incidence of the outcome). Discrimination measures how well a prognostic model can differentiate subjects with and without outcome. This aspect of model performance can be assessed using the c-statistic defined as the area under the ROC curve (AUC). An AUC of 1 indicates perfect discrimination whereas an AUC of 0.5 indicates that the model does not discriminate better than random guessing. Calibration measures the agreement between observed and predicted outcomes. This measure can be quantified in terms of calibration-in-the-large and the calibration slope. For binary outcome, calibration-in-the-large was estimated as the intercept of the logistic regression model that regresses observed outcomes on the linear predictor derived from the prediction model with its slope forced to be 1 (i.e. the linear predictor is included as an offset). Therefore, this measure assesses how well the predicted risks match the observed outcomes in the log-odds scale, adjusted for the linear predictor. The optimal value of calibration-in-the-large is 0. Calibration-in-the-large of <0 or >0, respectively, indicate that predicted outcomes are systematically too high or too low. The calibration slope was estimated as the slope of the logistic regression model that regresses observed outcome on the linear predictor derived from prediction model. This measure reflects the extremeness of the predicted outcome and is compared to 1. A calibration slope < 1 indicates that the predictions are too extreme; whereas a calibration slope > 1 implies that the predictions are not extreme enough.

Correction for optimism As no independent validation datasets were available to assess performance of prediction models developed in this project, models were evaluated on the development dataset which imposes the risk of optimism, i.e. over-estimation of performance due to over-fitting (Steyerberg, 2010). In order to compensate for optimism and to get a realistic assessment of the performance of the entire model development process, all performance measures were corrected for optimism using temporal validation or 10-times

repeated 10-fold cross-validation technique.

In temporal validation, the whole data was split into a training and a test set based on time. The whole model development process was applied on the training set which was the old data, and subsequently, the derived model was assessed on the test set which was the new data. This so-called "temporal performance" represents a more truthful assessment of the whole modeling process than the optimistic apparent one (Steyerberg, 2010) .

In 10-times repeated 10-fold cross-validation, the whole modeling procedure except for assumption assessments was firstly repeated 10 times by using a selection of nine tenths of the data for model development and one tenth for validation, respectively (Steyerberg, 2010). The cross-validation was further repeated ten times to minimize dependence on the random split into ten sub-datasets. The performance of the derived model on the test sets was then averaged across the 100 test sets to provide overall optimism-corrected performance measures.

Model presentation

The final model which had the best trade-off between simplicity and accuracy was chosen as the basis for a score chart following the approach of Sullivan et al. (2004). In brief, the linear predictor of the selected model was rounded and simplified, followed by a categorization of continuous variables and assignment of a point value to each category of a covariate. The total point score for each patient obtained from the score chart is an approximation of the linear predictor corresponding to that patient which can then be converted to a predicted risk. Finally, the adequacy of this score chart was evaluated by comparing risk predictions from the score chart to those of the original statistical model, and by visualizing their agreement with a Bland-Altman plot (Bland and Altman, 1986).

Adjustment of model development and validation steps for multiple imputation

For multiple imputed datasets, the above model development and validation steps were adjusted according to current recommendation (White et al., 2011).

For assessing the linearity and additivity assumptions, likelihood ratio test were performed comparing the simple model with linear terms and no interaction terms to models with more flexible terms for each imputed dataset. The results were then combined with

method given by Meng and Rubin (1992). Likewise, the overall multivariable model was derived by using Rubin's rule to combine the multivariable models fitted on each imputed dataset.

Variable selection was based on backwards stepwise model selection (Hastie et al., 2009). At each variable selection step, the model of interest was fitted to all imputed datasets and the least significant predictor was excluded if its pooled p value was larger than 0.15 (the p-value cut-off of 0.15 was chosen to approximately mimic variable selection based on AIC). The final model was obtained by applying Rubin's rule to aggregate parameter estimates from a model which included all predictors that remained after the variable selection procedure across imputed datasets.

Regarding model validation, both temporal and cross-validation were performed as described above. In both cases, the whole modeling procedure for each statistical model of interest was applied to each imputed training set. Predictions of each fitted model on the corresponding imputed test set were obtained and then compared to observed outcomes in the test set of each imputed dataset to derive performance measures. These measures were then averaged across imputed test sets to provide a single set of measures for each model.

2.2.4 Statistical software

All analyses were performed with the statistical software R version 3.1.2 (2014-10-31) (R Core Team, 2014) and its companion packages including machine learning and multiple imputation packages (as described in the previous sections) and other packages including Hmisc version 3.14.6 (Harrell and with contributions from Charles Dupont and many others, 2014), ggplot2 version 1.0.0 (Wickham, 2009), plyr version 1.8.1 (Wickham, 2011) and dplyr version 0.3.0.2 (Wickham and Francois, 2014).

Chapter 3

Clinical and laboratory features of children with DSS

Summary

DSS is a severe manifestation of dengue virus infection that particularly affects children and young adults. Despite its increasing global importance, there are no prospective studies describing the clinical characteristics, management or outcomes of DSS. This chapter describes the findings at onset of shock and the clinical evolution until discharge or death, based on a comprehensive prospective dataset of 1719 Vietnamese children with laboratory-confirmed DSS managed in a single intensive care unit between 1999 and 2009.

The research in this chapter has been published in: Lam PK, Tam DTH, Diet TV, Tam CT, Tien NTH, Kieu NTT, Simmons C, Farrar J, Nga NTN, Qui PT, Dung NM, Wolbers M, Wills B (2013) Clinical characteristics of dengue shock syndrome in Vietnamese children: a 10-year prospective study in a single hospital. *Clinical Infectious Diseases*, 57(11): 1577–86.

3.1 Introduction

Despite the increasing burden of dengue globally and the severity of DSS which is potentially fatal, only a few small retrospective reports have described the clinical characteristics, management, and outcomes of DSS cases (Bunnag and Kalayanarooj, 2011; Marón et al., 2011; Ranjit et al., 2005). At the HTD in Ho Chi Minh City a prospective observational study aiming to enrol all children presenting with DSS was conducted between 1999 and 2009. This chapter presents data from more than 1700 cases collected over the 10-year period, providing the first comprehensive description of the clinical features of DSS in children.

3.2 Methods

In this chapter, data from patients enrolled into the DF cohort were used to describe the clinical features and outcome of DSS in Vietnamese children. Details regarding study design and data collection, including dengue confirmation and serological classification, were described in Chapter 2. Disease classification was performed using the WHO 1997 and 2009 criteria (World Health Organization, 1997, 2009). The total number of DSS cases admitted directly to the PICU during the study period (excluding transfers from other hospitals after initiation of resuscitation) was ascertained from the hospital's main record system.

3.3 Results

From 1999-2009 a total of 1810/1847 children (98%) admitted to PICU with clinical DSS participated in the study. In 19 cases both RT-PCR and paired serology were negative, while in 72 cases the results were inconclusive; in the remaining 1719 cases (95%) dengue virus infection was confirmed, with the infecting serotype identified in 1209/1647 cases (73%) for whom RT-PCR was performed. Among the confirmed dengue patients 503 (29%) participated in the nested RCT (as described in Chapter 2), while the remaining 1216 (71%) were enrolled in the observational study. Almost all cases were admissions from the local catchment area, with less than 4% of cases transferred from another health

facility, most during the febrile phase; however two patients were enrolled in error, having already received parenteral fluid therapy for shock resuscitation prior to transfer.

3.3.1 Characteristics at presentation with shock

Demographic information and selected clinical characteristics for all 1719 confirmed dengue patients are described in Table 3.1. For most study-specific parameters data was missing in less than 5% of cases. The median age was 10 years, varying by year during the study from 9 to 11 years. The median (IQR) day of illness at shock was consistently 5 (4-6) for each year of the study, although 62 cases (4%) overall presented on illness day 3.

The most common symptoms reported were lethargy (1490/1719, 87%), vomiting (1199/1713, 70%) and abdominal pain (1238/1714, 72%). Most children were afebrile, but 153/1718 (9%) still had an axillary temperature of 38°C or more at onset of shock, without a clear relationship to the day of illness at that time (p=0.1, Wilcoxon rank-sum test). In 123/1719 (7%) the blood pressure was not measureable, while 417/1596 (26%) of the remainder exhibited hypotension for age and 1568/1596 (98%) had a pulse pressure of 20 mmHg or less. Respiratory distress (3/1718, <1%) and cyanosis due to profound shock (10/1714, <1%) were extremely uncommon. The liver was palpable in 1478/1696 (87%) of cases, with abdominal tenderness in 1238/1714 (72%), whereas a palpable spleen was extremely uncommon (only 5 cases documented). Almost one third (493/1719, 29%) of the patients had no evidence of bleeding. Among cases with bleeding this was limited to skin petechiae or minor bruising in the majority of cases, with mucosal haemorrhage noted in only 73 cases.

3.3.2 Progress in hospital

Since many patients in the RCT received initial resuscitation with a colloid according to their randomization, information on management and complications after enrolment is presented for the observational study and RCT groups separately (Table 3.2). Apart from the greater colloid usage there was little difference between the two study groups other than a slightly higher proportion of minor skin bleeding observed in the RCT group. Considering the observational study only, most children recovered well with standard crystalloid resuscitation, although 547/1211 (45%) patients also received colloid therapy, 244

Table 3.1. Baseline characteristics of the study participants at enrolment (n = 1719).

Characteristics	Observational study (n = 1216)			All patients (n = 1719)		
	n	Summary statistics		n	Summary statistics	
Age [year]	1216	10	(7-12)	1719	10	(7-12)
Gender: Female	1216	567	(47)	1719	817	(48)
Referal status	1216			1719		
- Home		477	(39)		720	(42)
- HTD		673	(55)		911	(53)
- Other		58	(5)		65	(4)
- Unknown		8	(1)		23	(1)
Day of illness	1216	5	(5-6)	1719	5	(4-6)
Weight [kg]	1216	29	(21-38)	1719	27	(20-35)
Temperature $\geq$ 38°C	1215	112	(9)	1718	153	(9)
Pulse rate [per min] [a]	976	120	(104-120)	1393	120	(100-120)
Systolic BP [mmHg] [a]	1138	90	(85-100)	1596	90	(85-100)
Pulse pressure [mmHg] [a]	1138	20	(15-20)	1596	20	(15-20)
Haemorrhage	1216			1719		
- None		402	(33)		493	(29)
- Skin only		774	(64)		1153	(67)
- Mucosal		40	(3)		73	(4)
Abdominal tenderness	1214	794	(65)	1714	1238	(72)
Liver size [cm]	1204	2	(1-2)	1696	2	(1-2)
Haematocrit [%]	1195	50	(47-52)	1696	49	(46-52)
Platelet count [1,000 cell/mm^3]	1196	38	(26-54)	1695	41	(28-61)
AST [IU/l]	917	133	(89-218)	1030	125	(80-206)
DHF according to WHO 1997	1159	635	(55)	1642	939	(57)
RT-PCR performed	1176			1647		
- DENV-1		661	(56)		675	(41)
- DENV-2		285	(24)		367	(22)
- DENV-3		19	(2)		48	(3)
- DENV-4		8	(1)		110	(7)
- Mixed		8	(1)		9	(1)
- Negative		195	(17)		438	(27)
Immune status	1115			1618		
- Primary		6	(1)		6	(0)
- Secondary		1024	(92)		1506	(93)
- Unclassifiable		85	(8)		106	(7)

Summary statistics are median (IQR) for continuous variables and frequency (%) for categorical variables.

[a] *Only for subjects with measurable value.*

Abbreviation: IQR = interquartile range, HTD = Hospital for Tropical Diseases, BP = blood pressure, AST = aspartate aminotransferase, DHF = dengue haemorrhagic fever, WHO = World Health Organization, RT-PCR = reverse transcriptase polymerase chain reaction, DENV = dengue virus.

(45%) of them within the first 2 hours. Most children (328, 60%) in this group received only a single colloid bolus, but up to 7 colloid infusions were needed for severe cases, with a median (IQR) volume of 19 (12-25) ml/kg of colloid given throughout hospitalization, on a background of 114 (99-129) ml/kg total parenteral fluid therapy. Considering the whole patient cohort, additional cardiovascular support with inotropic drugs was required in 75/1719 (4%), and 513/1717 (30%) patients developed clinical signs of fluid overload (pleural effusion or ascites) following resuscitation. Among these patients, 313/513 (61%) were treated with diuretic therapy for 1-2 days after haemodynamic stabilization.

After admission 158/1719 (9%) children developed at least one new bleeding manifestation, among them 98 cases with skin bleeding only and 60 cases with mucosal bleeding. Considering all 126 patients with overt mucosal bleeding (either present at enrolment or developing subsequently) gastrointestinal bleeding occurred most frequently (61 cases), compared to epistaxis (36 cases), gum bleeding (22 cases) or unusual vaginal bleeding (21 cases). In 31 patients overall the bleeding was clinically severe, requiring transfusion in 26 cases (18 during active resuscitation, and 8 during the recovery phase due to symptomatic anaemia), resulting in significant but asymptomatic anaemia at discharge in 4 cases, and involving a critical organ in 1 case (spinal cord haemorrhage, confirmed on magnetic resonance imaging (MRI) scan). Although most severe haemorrhage involved the gastrointestinal tract primarily (15 cases), 7 children had severe skin bleeding only, mainly at sites of invasive procedures, and 4/7 required transfusion. Platelet concentrates were not available during the study but children with severe coagulopathy and bleeding received fresh frozen plasma or other blood products at the discretion of the treating clinician.

Table 3.2. Summary of complications, management and outcomes during hospitalization (n = 1719).

Characteristics	Observational study (n = 1216)			RCT (n = 503)			All patients (n = 1719)		
	n	Summary statistics		n	Summary statistics		n	Summary statistics	
New bleeding	1216	77	(6)	503	81	(16)	1719	158	(9)
Site of new bleeding	1216			503			1719		
- None		1139	(94)		422	(84)		1561	(91)
- Skin only		39	(3)		59	(12)		98	(6)
- Mucosal		38	(3)		22	(4)		60	(3)
Severe bleeding	1216	20	(2)	503	11	(2)	1719	31	(2)
Transfusion	1216	18	(1)	503	8	(2)	1719	26	(2)
Inotropic drug	1216	55	(5)	503	20	(4)	1719	75	(4)
Total parenteral fluid volume [ml/kg]	1210	114	(99-129)	503	125	(110-143)	1713	116	(102-133)
Used colloid	1211	547	(45)	503	418	(83)	1714	965	(56)
Total colloid volume [ml/kg][a]	546	19	(12-25)	418	25	(25-35)	964	25	(15-29)
Clinical fluid overload	1214	340	(28)	503	173	(34)	1717	513	(30)
DHF according to WHO 1997	1206	796	(66)	499	406	(81)	1705	1202	(70)
Survival status: Died	1216	7	(1)	503	1	(<1)	1719	8	(<1)

Summary statistics are median (IQR) for continuous variables and frequency (%) for categorical variables.

[a] *Only for subjects received colloid.*

Abbreviations: IQR = interquartile range, RCT = randomised controlled trial, DHF = dengue haemorrhagic fever, WHO = World Health Organization.

The evolution of haematocrit and platelet values during hospitalization is shown in Figure 3.1. The median (IQR) maximum haematocrit was 50% (47-52), documented at presentation in most cases (86%, 1484/1719). Among cases with both enrolment and one-month follow-up haematocrit values 753/830 (91%) had evidence of at least 20% haemoconcentration at enrolment. The haematocrit declined rapidly during the first 4 hours of fluid resuscitation, later rising again in the majority of children. In contrast the platelet nadir (median (IQR) of 28,000 (19,000-40,000 cells/mm^3) occurred most frequently one day after onset of shock (720/1718, 42%). Although a transient drop in platelet count was seen in all cases, in 25/1718 (1.5%) cases the nadir did not fall below 100,000 cells/mm^3. Coagulation profiles were performed infrequently and are not reported here, but the abnormalities observed were consistent with previous reports (Wills et al., 2002, 2009). Liver enzyme levels were checked in approximately 60% and were moderately elevated at shock, with aspartate aminotransferase levels consistently higher than alanine aminotransferase levels.

All patients would have fulfilled the 2009 WHO criteria for severe dengue, while only 939/1642 (57%) of the children with sufficient data to allow classification at the onset of shock would have been categorized as DHF. Using all available information from the acute illness and any follow-up visits, 1202/1705 (70%) of the patients eventually fulfilled the four criteria for DHF with the remainder classified as dengue fever by default.

3.3.3 Outcome

The most common complication observed during treatment for DSS is recurrent shock, conventionally termed "re-shock"; the accepted definition at HTD is narrowing of the pulse pressure (PP) to $\leq$ 20 mmHg after a period of apparent cardiovascular stability, associated with tachycardia and cool extremities, and considered to require additional volume resuscitation with a colloid fluid bolus. Patients may experience several episodes of re-shock during the critical period for leakage. According to local management guidelines the need for two or more colloid boluses (given either at presentation with decompensated/hypotensive shock or during re-shock episodes) is considered an indicator of severe disease, and is the recommended threshold to proceed to central venous pressure (CVP) monitoring. In cases with ongoing hypotension and a poor response to colloid therapy,

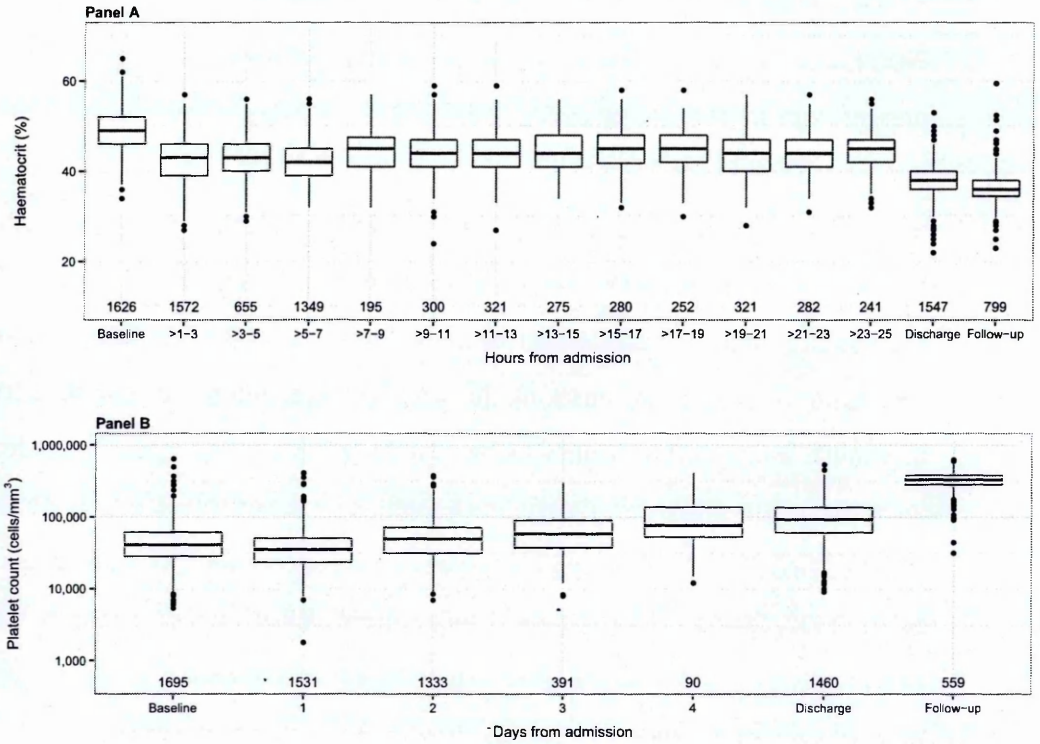

Figure 3.1. *Boxplots describing changes in haematocrit (upper panel) and platelet count (lower panel) during the evolution of the illness. Haematocrit data is presented for the 24 hours following admission, while platelet data is presented daily for the first 4 days, together with the values at hospital discharge and follow-up for both parameters. The numbers displayed below each boxplot represent the number of patients included within that time interval. If multiple values were recorded during any time interval, the highest haematocrit and the lowest platelet count were respectively chosen for that patient. The haematocrit graph excludes data from the 73 DSS cases with mucosal bleeding at presentation.*

inotropic agents such as dopamine or dobutamine may be added. Other major complications include severe bleeding (requiring a blood transfusion, involving a critical organ, or resulting in significant but compensated anaemia), and organ failure (significant impairment in function of an organ system).

These facts suggested three main outcomes for consideration among children with DSS. The first potential outcome is "recurrent shock" which was defined as development of one or more episodes of re-shock after the initial resuscitation. The second outcome was "critical DSS", here defined as death or requirement for inotropes (in addition to colloid therapy to maintain cardiovascular stability) or development of any major complication (severe bleeding or organ failure). The last main outcome was a composite outcome of

"profound DSS", defined as either a) 2 or more episodes of re-shock in subjects presenting with compensated shock, or b) 1 or more episodes of re-shock in subjects presenting with decompensated/hypotensive shock (thus these patients had already received a colloid bolus during their initial resuscitation), or c) requirement for inotropes or development of any other major complication, or d) death. (In the first two categories the participants would have been treated with at least two colloid boluses, i.e. they would have achieved the threshold for CVP monitoring).

To assess whether a patient had recurrent shock or profound DSS, detailed information on fluid resuscitation of each patient is required. Information on fluid use was missing in 10 cases, but among the 1709 patients with complete information on fluid usage, 595 (35%) developed recurrent shock at some point after the initial resuscitation, and 367 (21%) had profound DSS. Amongst all patients, 86/1719 (5%) fulfilled the criteria for critical DSS.

Looking specifically at mortality, only 8 patients died during the 10-year study period, including 1 infant and 7 children (Table 3.3), although one additional DSS-associated death outside the study was identified from hospital records. In 3/8 cases shock occurred early, on illness day 4. All 8 patients developed profound shock within the first 12 hours, requiring multiple colloid infusions plus inotropic support and with rapid development of significant fluid overload. The interval from admission to death was generally short (median, range 34 (11-87) hours in 7 cases) and one child with multi-organ failure was taken home moribund after 4 days. Major bleeding requiring transfusion was apparent in 7/8 cases before death.

Overt organ dysfunction was very uncommon. Other than in association with prolonged shock no patient in the cohort had clinically significant hepatic, renal or neurological compromise, except for the child with spinal cord haemorrhage and one other child with profound shock, liver failure and coma. The latter two children gradually improved over several weeks with supportive care, and both eventually made a full recovery.

Table 3.3. Selected clinical and laboratory characteristics for the 8 children who died.

Characteristics	Case 1 [a]	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8 [b]
At presentation with shock								
Age [months (m) or years (y)]	11m	7y	6y	7y	7y	13y	6y	3y
Gender	Male	Male	Female	Female	Male	Female	Male	Male
Year of study	2000	2006	2007	2007	2007	2009	2009	2009
Day of illness [days]	5	4	5	5	5	6	4	4
Temperature [°C]	39.0	37.0	38.0	39.2	37.5	37.0	37.0	37.0
Pulse rate [per min]	120	Rapid, weak	120	140	120	152	Rapid, weak	Rapid, weak
Systolic BP [mmHg]	90	95	85	90	85	95	0	90
Pulse pressure [mmHg]	10	15	25	20	10	15	0	20
Bleeding	Petechiae	Petechiae	Petechiae, GI	-	-	Petechiae	Petechiae	-
Abdominal tenderness	+	+	+	+	+	+	-	-
Liver size [cm]	4	2	3	2	2	3	1	-
HCT [%]	41	N/A	N/A	42	51	56	53	53
PLT [cell/mm^3]	108,000	7,000	35,000	38,400	N/A	39,000	55,800	97,000
During hospitalization								
Maximum HCT [%]	41	53	53	46	51	56	53	53
Minimum PLT [cells/mm^3]	108,000	7,000	35,000	13,300	30,400	20,000	7,000	17,700
New bleeding	GI	-	-	GI	GI	GI	-	GI, epistaxis
Transfusion	+	-	+	+	+	+	+	+
Inotropic drug	+	+	+	+	+	+	+	+
Number of colloid boluses	2	2	2	3	5	3	5	5
Total colloid volume [ml/kg]	35.8	23.5	11.0	33.1	85.2	54.1	107.5	90.2
Overload	+	+	+	+	+	+	+	+
Hours from admission to death	11	23	24	39	34	37	87	N/A
Serotype	DENV-3	DENV-1	DENV-1	DENV-3	DENV-3	DENV-1	DENV-1	DENV-1
Serology	Possible primary	Secondary	Secondary	Secondary	Secondary	Secondary	Secondary	Secondary

[a] *Case 1 was enrolled in the fluid randomised controlled trial and received the first bolus of colloid according to the trial randomization.*

[b] *Case 8 was taken home 4 days after admission and is presumed to have died. The child was profoundly hypotensive with multiorgan failure at the time of discharge.*

Abbreviations: BP = blood pressure, HCT = haematocrit, PLT = platelet count, DENV = dengue virus, GI = gastrointestinal bleeding, + = yes, - = no, N/A = not available.

3.3.4 Dengue serotypes and immune status

The relative frequency of dengue serotypes identified in the patient cohort over time is presented in Figure 3.2, Panel A. With increasingly sensitive diagnostics the proportion of cases with a serotype identified increased gradually, rising from 33% initially to more than 82% after 2007. In 1999 DENV-3 was the most common serotype seen, replaced by DENV-4 peaking in 2001, DENV-2 peaking in 2004, and finally by DENV-1 extending from 2005 to 2009. Almost all patients had an IgG response consistent with secondary infection, although in 106/1618 (7%) of cases the information available was insufficient to allow categorization. The pattern of serotypes observed in the children with DSS was very similar to that seen among 1509 children with secondary dengue without shock enrolled into the MD cohort during 2001-2009 (Figure 3.2, Panel B).

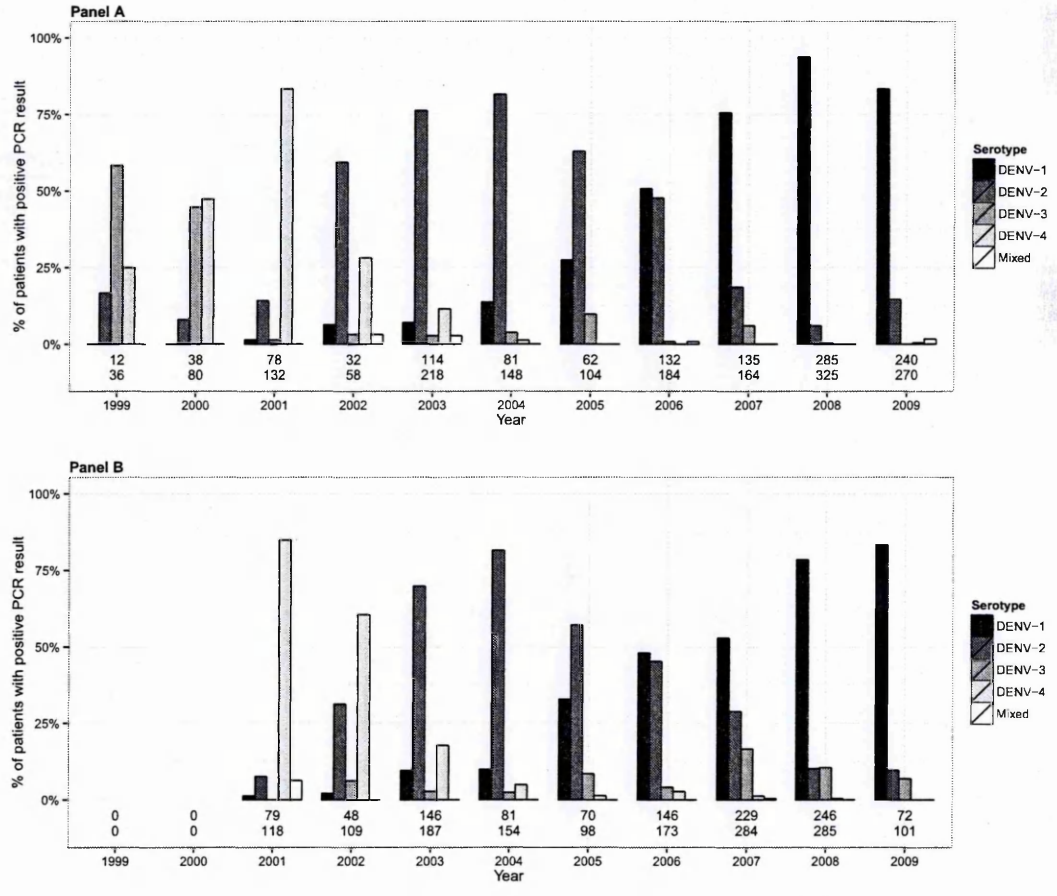

Figure 3.2. *Serotype distributions over time for DSS cases (Panel A), and for children with secondary dengue but did not experience severe complications in the MD cohort (Panel B). The numbers below each bar are the total number of cases in whom a serotype was identified (first line), and the total number of cases enrolled into the corresponding study (second line).*

Table 3.4. *Selected clinical and laboratory characteristics for the 6 primary dengue cases.*

Characteristics	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
At presentation with shock						
Age [months/years]	4m	11m	13m	7y	1y	12y
Gender	Male	Female	Male	Male	Female	Male
Year of study	2008	2008	2008	2008	2009	2009
Day of illness	5	5	5	5	5	6
Temperature [°C]	37.5	37.0	37.0	37.0	37.0	37.0
Pulse rate [per min]	148	Rapid, weak	168	140	Rapid, weak	110
Systolic BP [mmHg]	50	70	90	90	80	90
Pulse pressure [mmHg]	20	20	15	20	20	20
Bleeding	Petechiae	Petechiae	Petechiae	-	Petechiae	Petechiae
Abdominal tenderness	+	+	+	+	-	-
Liver size [cm]	3	1	4	3	2	1
HCT [%]	36	47	50	56	45	46
PLT [cell/mm^3]	40,000	19,000	77,000	81,400	37,900	66,000
During hospitalization						
Maximum HCT [%]	36	47	50	56	45	47
Minimum PLT [cells/mm^3]	24,000	9,000	45,800	61,000	28,400	53,000
New bleeding	-	-	-	-	-	-
Number of colloid boluses	0	2	1	0	0	0
Colloid volume [ml/kg]	0.0	25.0	25.3	0.0	0.0	0.0
Clinical fluid overload	-	-	+	-	-	-
Survival status	Recovery	Recovery	Recovery	Recovery	Recovery	Recovery
IgG by day of illness	Day 5: (-)	Day 5: (-)	Day 5: (-)	Day 5: (-)	Day 5: (-)	Day 6: (-)
	Day 8: (-)	Day 8: (-)	Day 8: (-)	Day 8: (-)	Day 8: (-)	Day 9: (-)
Serotype	DENV-2	DENV-1	Negative	DENV-1	DENV-1	DENV-1

Diagnosis for the case with PCR negative was based on the positive dengue IgM capture ELISA on samples taken on days 5 and 8.

Abbreviations: BP = blood pressure, HCT = haematocrit, PLT = platelet count, IgG = immunoglobulin G, IgM = immunoglobulin M, ELISA = enzyme-linked immunosorbent assay, DENV = dengue virus, + = yes, - = no.

Overall only 6 cases were classified as clear primary infections, 4 infants and two children aged 7 and 12 years (Table 3.4). For an additional 5 children under 18 months immune status was classified as indeterminate, but the serological patterns observed and their age suggested primary infection. Conversely, all 157 children aged 18-60 months with classifiable immune status had secondary dengue. All the definite primary cases recovered, although two infants required colloid infusions. However, one 11-month old boy with indeterminate/possible primary dengue died with profound shock and major gastrointestinal bleeding.

3.4 Discussion

This chapter presents the first comprehensive description of the clinical presentation of DSS in children, using data gathered prospectively over 10 years on a large cohort of patients managed in a single Vietnamese institution. Over 95% of all children admitted with DSS during the study period were evaluated. Since prior shock resuscitation might confound the clinical picture the analysis focused on direct admissions only; although a few cases were missed, including one child who died, overall the results are representative of the clinical spectrum of DSS cases admitted directly to a busy hospital in a hyperendemic region.

During the 10-year study each dengue serotype predominated for one or more years, so DSS caused by all four serotypes were able to be observed. Apart from infants below 18 months, virtually all children had secondary dengue, in line with established concepts of pathogenesis (Yacoub et al., 2013). The pattern of serotype replacement seen in the cohort was similar to that seen among children with secondary infections enrolled the MD cohort, and also to the relative virus prevalence identified by passive surveillance in southern Vietnam during the same time-period (Vu et al., 2010). Thus the viruses associated with DSS appear to be representative of the virus population affecting the wider community with no evidence that a particular serotype contributes to a greater risk for shock. Notably however 3 out of 8 deaths were associated with DENV-3 although the total number of DENV-3 infections identified was small. Since a number of interacting host and viral factors influence an individual's propensity to develop severe vascular leakage (Yacoub et al., 2013), only very detailed studies can establish whether particular viral characteristics do confer an increased risk for DSS or death.

The clinical signs and symptoms documented in this large cohort were generally consistent with empirical descriptions of DSS (World Health Organization, 2009). Interestingly however, 9% of all cases were still febrile at presentation. Increased permeability commences during the febrile phase, typically resulting in shock when leakage exceeds the capacity of the homeostatic compensatory mechanisms to maintain adequate plasma volume (Srikiatkhachorn et al., 2007; Trung and Wills, 2010). Functional cardiac impairment also contributes to the cardiovascular decompensation, although the underlying

mechanisms remain unclear (Yacoub et al., 2012). Although defervescence and onset of DSS are often temporally linked it is important that clinicians managing early dengue cases are aware that DSS can occur before defervescence. Identification of more reliable warning signs of likely deterioration would be useful both for individual case-management and to facilitate effective use of limited healthcare resources.

In agreement with other studies (Phuong et al., 2004), a considerable number of DSS cases had no bleeding manifestations during the illness episode. Severe bleeding was uncommon and primarily observed from the gastrointestinal tract, although massive soft-tissue bleeding necessitating transfusion occurred in 3 children. Also consistent with other studies (Alexander et al., 2011; Phuong et al., 2004), almost one third of cases did not achieve the WHO 1997 classification for DHF, mainly due to failure to fulfill the haemorrhage and/or plasma leakage criteria since thrombocytopenia was almost universal. All patients were examined carefully each day but a tourniquet test was not mandatory as these are infrequently performed in Vietnam and several studies have demonstrated poor utility in clinical practice (Phuong et al., 2004; Srikiatkhachorn and Green, 2010). Moreover, radiological investigations to identify plasma leakage were not performed unless clinically indicated, as the study aimed to reflect real-world practice. Haemoconcentration below the conventional threshold of 20% in association with DSS has been reported previously (Marón et al., 2011), and radiological evidence of leakage is often not detected until relatively late in the disease evolution. Since patients must be treated according to their actual clinical status at any time, it is apparent that the 2009 WHO classification system is preferable for individual case-management (World Health Organization, 2009).

The case fatality rate was extremely low (<1%). Most patients recovered well with the standard crystalloid fluid regimen or following a single bolus of colloid, and requirement for additional colloid therapy, inotropic support and/or blood products was infrequent. Prompt diagnosis and immediate admission to PICU with management coordinated by a highly experienced team undoubtedly contributed to this favourable outcome. In line with WHO principles for fluid management of DSS, the unit operates a generally conservative policy after initial resuscitation, relying on frequent clinical assessments and regular ward-based haematocrit measurements to limit fluid administration to the minimum required, thereby minimizing the risk of fluid overload. However, the study focused on direct ad-

missions only and it is clear that external referrals with prolonged shock or established fluid overload are considerably more difficult to manage and have correspondingly higher mortality rates (Bunnag and Kalayanarooj, 2011).

Only a very small number of confirmed primary dengue cases were included in the cohort and all recovered quickly without notable complications. However one death did occur in a suspected primary case, underlining the view that primary dengue can result in severe and even fatal disease (Barnes and Rosen, 1974; Nogueira et al., 1999; Scott et al., 1976). Given that immune status could not be defined in 6% of patients some primary cases might have been missed but the number is likely to be small.

Three potential outcomes for patients with DSS were identified as potentially useful for developing prognostic models in this population, including recurrent shock, profound DSS and critical DSS. However, profound DSS is preferred as the primary outcome of interest for a number of reasons. Firstly, the number of subjects experiencing critical DSS was too small for prognostic modelling. Secondly, differences in initial resuscitation between patients with compensated or hypotensive shock (use of crystalloid or colloid fluids, respectively) may have influenced the likelihood of developing recurrent shock subsequently. Thirdly, a large proportion of cases experiencing their first episode of re-shock recover fully following a single colloid bolus without requiring additional supportive therapy. By defining as a composite outcome measure that includes use of at least two boluses of colloid, profound DSS is able to reflect the local threshold for concern regarding severe disease, as indicated by the recommendation to proceed to CVP monitoring in cases requiring more than two colloid boluses.

In summary, this is a comprehensive clinical description of DSS in a large cohort of Vietnamese children. With prompt intervention and assiduous clinical care by experienced staff the outcome of this potentially fatal condition can be very good. As the emerging dengue pandemic spreads to new geographical locations it is crucial that this accumulated experience be translated into practical advice and support for clinicians newly exposed to this severe complication of a common disorder.

Chapter 4

Prognostic models for profound DSS amongst children with DSS

Summary

Reliable prognostic tools to assist physicians in identifying children at risk of profound DSS and likely to require intensive support are lacking. This chapter used data from the DF cohort to identify clinical and laboratory risk factors of profound DSS, develop a prediction model for profound DSS and derived a simple score chart for use in clinical practice.

4.1 Introduction

Prognostic models for poor outcomes can enhance a physician's clinical decision-making processes (Steyerberg et al., 2013; Riley et al., 2013). Several prognostic models (PRISM, PIM) have been developed to characterise children with severe illness admitted to western PICUs (Pollack et al., 1988; Shann et al., 1997), and to compare outcomes within and between different units over time. However, models such as PRISM and PIM typically require detailed clinical and laboratory data that is not readily available in countries where dengue is endemic, and are thus not practical in these settings. Although a number of predictive models have been developed to help distinguish dengue from other febrile illnesses with similar presentations (Potts and Rothman, 2008; Potts et al., 2010a), and to try to improve identification of cases likely to develop severe complications (Tanner et al., 2008; Potts et al., 2010a), to date only a single report describes a prognostic model for poor outcome in patients with established DSS (Huy et al., 2013a). Therefore this study analyzed data from a large 10-year cohort of children presenting with DSS, aiming to identify risk factors for profound shock, and to develop a prognostic model to assist physicians in identifying children likely to require intensive supportive therapy.

4.2 Methods

This chapter utilized data from the DF cohort. Detailed information related to study design, study participants, dengue diagnostics, general statistical analyses (descriptive analysis, treatment of missing values) and modeling strategy are described in Chapter 2. I present here definitions of clinical outcomes and candidate predictors and several specific statistical methods used in this chapter.

4.2.1 Clinical outcomes and candidate predictors

As discussed in Chapter 3 (Section 3.4), amongst three potential clinical outcomes in children with DSS (recurrent shock, profound DSS, critical DSS), profound DSS is preferred as the primary outcome. Secondary outcomes were recurrent shock, critical DSS, and the total volume of colloid, defined as the total volume of colloid patient received during

hospitalization from shock. All clinical definitions are described in Table 4.1.

Table 4.1. *Definition of clinical outcomes amongst children with DSS.*

Clinical outcome	Definition
Profound DSS	Death OR major complications OR requirement of specific additional therapy (2 or more colloid boluses, or inotropic support)
Recurrent shock	Pulse pressure $\leq$ 20 mmHg after a period of apparent cardiovascular stability, associated with tachycardia and cool extremities, and considered to require additional volume resuscitation with a colloid fluid bolus
Critical DSS	Death OR major complications OR requirement of inotropic support
Total volume of colloid	The total volume of colloid (ml/kg) patient received during hospitalization from shock
Major complications	Severe bleeding OR organ failure
Severe bleeding	Requirement of blood transfusion OR bleeding resulting in significant but asymptomatic anaemia OR bleeding involving a critical organ
Organ failure	Significant impairment in function of an organ system, judged by the treating physician to require specific therapeutic intervention

Candidate predictors described in Table 4.2 were all assessed within 2 hours of onset of shock and were chosen based on clinical experience and evidence from the published literature (Srikiatkhachorn and Green, 2010; Wills et al., 2002; Marón et al., 2010). As pulse pressure (PP) and systolic blood pressure (BP) are closely linked haemodynamic parameters and some patients may present with no detectable blood pressure, an additional categorical candidate predictor, the haemodynamic index, was created to allow all patients to be classified into one of three ordered categories representing their initial cardiovascular status. The haemodynamic index is defined as 1 when the PP exceeds 10 mmHg and the systolic BP is maintained above the lower limit of normal for age (i.e. $\geq$ 80 mmHg if under 5 years, or $\geq$ 90 mmHg if aged 5 years or more). A haemodynamic index of 2 corresponds to a PP below 10 mmHg or a systolic BP below the lower limits for age, while a haemodynamic index of 3 indicates that the blood pressure is unmeasureable. Aspartate aminotransferase (AST) and dengue serotype were included in the univariate analysis but not in the multivariable analysis as they are less readily available for clinicians and were frequently missing (Table 4.3).

Table 4.2. List of candidate predictors.

Predictor	Unit or possible values	Type
Age	Year	Continuous
Gender	Female/Male	Binary
Weight	Kg	Continuous
Day of illness	Day of illness at shock	Continuous
Pulse rate	Beats per minute (fast and weak pulse = 200 pulses/min)	Continuous
Temperature	Body temperature [°C] measured in the axilla	Continuous
Systolic BP	mmHg	continuous
Pulse pressure	Difference between systolic and diastolic BP [mmHg] OR 5 mmHg if systolic BP was measurable but diastolic BP was unmeasurable	Continuous
Haemodynamic index	1 if systolic BP $\geq$ lower limit of normal[a] AND PP $\geq$ 10 mmHg 2 if systolic BP $<$ lower limit of normal[a] OR PP $<$ 10 mmHg 3 if systolic BP was unmeasurable	Categorical
Haemorrhage	"None" if no bleeding at enrolment "Skin" only if only have petechiae/bruising "Mucosal" if epistaxis OR gum OR gastrointestinal OR vaginal bleeding	Categorical
Abdominal tenderness	Yes/No	Binary
Liver size	Size of liver below costal margin [cm]	Continuous
HCT	Haematocrit value [%]	Continuous
PLT	Platelet count [cells/mm^3]	Continuous

[a] *Lower limit of normal systolic BP is 80 mmHg (if age $<$ 5 years old) or 90 mmHg (if age $\geq$ 5 years old).*
Abbreviations: BP = blood pressure, HCT = haematocrit, PLT = platelet count.

4.2.2 Statistical analysis

Analysis of profound DSS, recurrent shock and critical DSS

The study population included all patients with confirmed DSS for whom clinical outcomes could be assessed. Since the randomized treatment assignment might have affected the number of colloid boluses given, the primary analysis population excluded patients from the RCT (Wills et al., 2005). A sensitivity analysis including all patients was also performed, adjusting for the randomized treatment assignment by adding a categorical covariate with three levels (assigned to a colloid in the RCT; assigned to a crystalloid in the RCT; enrolled only in the observational study) to the corresponding regression models.

Details regarding the development of prognostic models are described in Chapter 2.

Logistic regression was the main statistical model for the univariate and multivariable analyses of all three outcomes (i.e. recurrent shock, profound DSS and critical DSS). Alternative statistical approaches were: logistic regression with variable selection and shrinkage based on the lasso, classification and regression trees (CART), generalized additive models (GAM), and gradient boosting with trees as base learners (Hastie et al., 2009).

To validate the modeling procedure, both temporal and internal validation of the whole model development process, except for the non-linearity and interaction assessments, were performed (Steyerberg, 2010). For temporal validation, the models were developed using data from the 939 patients enrolled before 2009 and validated on the 268 patients enrolled during 2009.

Analysis of total volume of colloid

As described in the Figure 4.1, the total volume of colloid had a very skewed distribution which consists of a point mass at 0 and a positive right-skewed tail. Therefore, normal linear regression cannot be applied and potential applicable analytical methods are robust regression (median regression) (Koenker, 2005), hurdle models (Tobin, 1958), or survival analysis. Survival analysis using Cox regression is traditionally used for survival analysis of time-to-event endpoints. However, it has also been suggested as a flexible model for general non-negative continuous outcomes with a right-skewed distribution and a point mass at 0 (Aalen et al., 2008). I decided to use the Cox proportional hazards regression model for the analysis of the total volume of colloid and treated patients who died either as right-censored observation or, alternatively, replaced their actual total colloid volume by the maximum observed total colloid volume plus 1.

4.3 Results

4.3.1 General description

A total of 1810 children were enrolled in the two studies (observational study and RCT) and the analysis population included 1706 patients with laboratory-confirmed dengue and complete information regarding fluid usage (Figure 4.2). Among these patients, the 1207

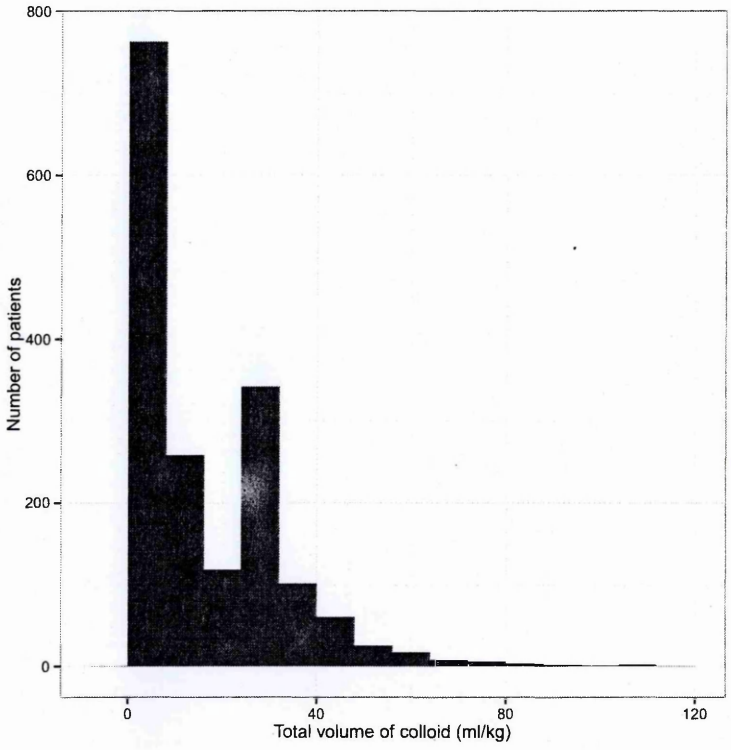

Figure 4.1. *Histogram of the total volume of colloid in all patients (n = 1706).*

patients enrolled in the observational study formed the primary analysis population.

A detailed description of patients enrolled into the DF study is provided in Chapter 3 and for reference, patient characteristics and outcomes are also reported in Table 4.3. In general, characteristics and outcome of patients enrolled only in the observational study were similar to all patients, except for the usage of colloid fluid which can be explained by the inclusion of patients enrolled into the fluid rescuscitation randomized clinical trial. In the main analysis population, 222/1207 (18%) of the children had profound DSS, 433/1207 (36%) developed recurrent shock, and 57/1207 (5%) had critical DSS. Deaths were very rare and major complications were infrequent; thus most children were classified as having profound DSS on the basis of their requirement of specific additional therapy. No systematic time trends were observed for the prevalence of profound DSS or critical DSS over the study period but there was a small but statistically significant decline in the prevalence of recurrent shock (linear trend tests: p values were 0.34, 0.03, 0.65 for profound DSS, recurrent shock and critical DSS, respectively) (Figure 4.3).

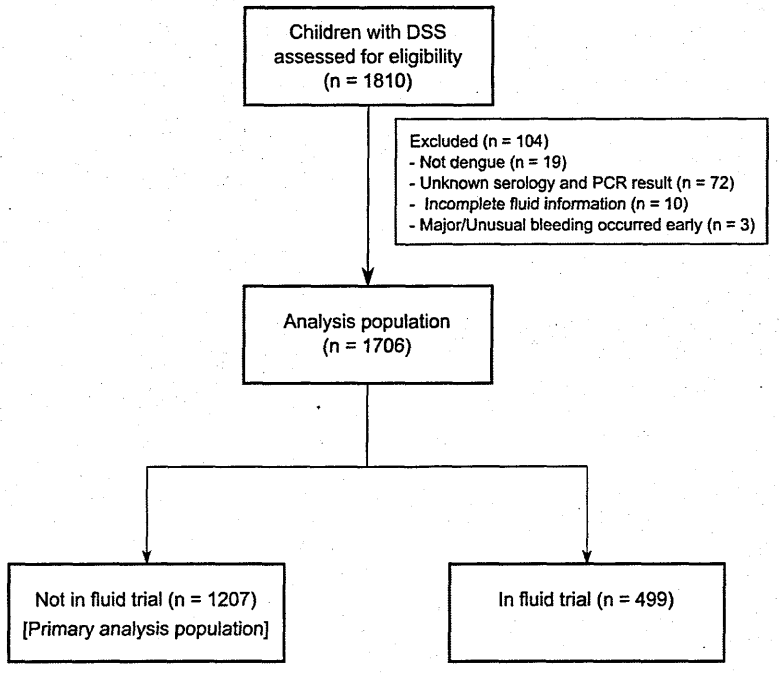

Figure 4.2. *Flow-chart of the analysis for profound DSS.*

Assessment of missing values Amongst all 1706 cases, 4% (75) of participants had at least one missing value in one or more candidate predictors. The number of missing values per individual ranged from 0 to 2. HCT and PLT were the two most frequently missing predictors with 1% missing values.

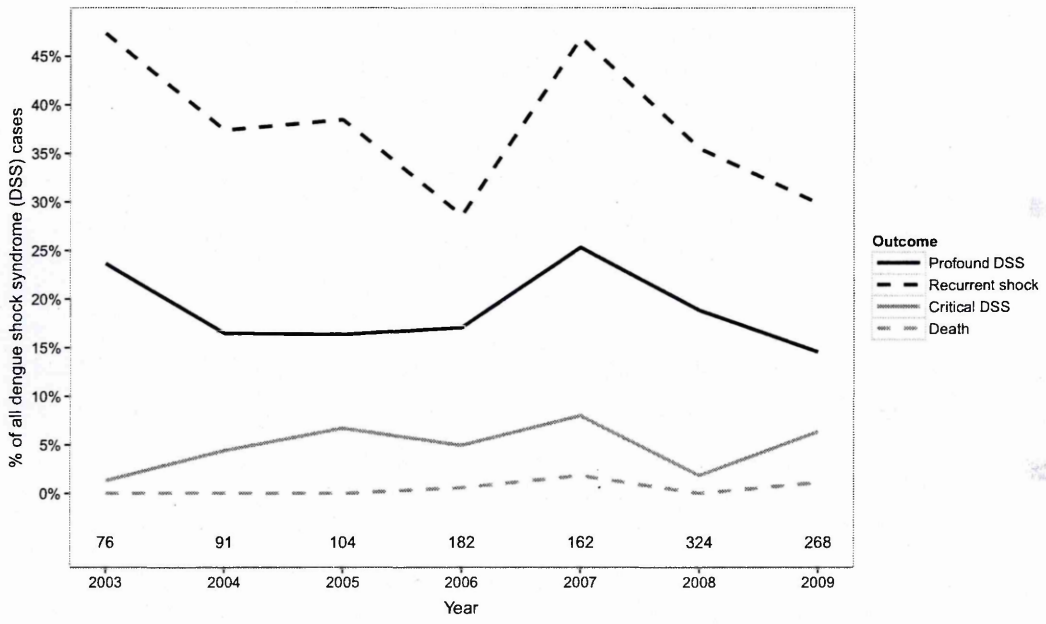

Figure 4.3. *Frequency of adverse clinical outcomes over time for patients enrolled in the observational study (n = 1207). The numbers shown below the line graphs indicate the total number of DSS cases enrolled in the study each year.*

Table 4.3. Baseline characteristics and outcomes of study participants in DF cohort.

Characteristics	Observational study (n = 1207)			All patients (n = 1706)		
	n	Summary statistics		n	Summary statistics	
Demographic characteristics						
Age [year]	1207	10	(7-12)	1706	10	(7-12)
Gender: Female	1207	562	(47)	1706	810	(47)
Clinical features at presentation with shock						
Weight [kg]	1207	29	(21-38)	1706	27	(20-35)
Day of illness [day]	1207	5	(5-6)	1706	5	(4-6)
Temperature $\geq$ 38°C	1206	108	(9)	1705	149	(9)
Pulse rate [per min]	1207	120	(100-140)	1706	120	(100-130)
Systolic BP [mmHg]	1207	90	(85-100)	1706	90	(80-100)
Pulse pressure [mmHg]	1207	20	(15-20)	1706	20	(15-20)
Haemodynamic index	1207			1706		
- Group 1		829	(69)		1146	(67)
- Group 2		300	(25)		438	(26)
- Group 3		78	(6)		122	(7)
Haemorrhage	1207			1706		
- None		398	(33)		489	(29)
- Skin only		769	(64)		1144	(67)
- Mucose		40	(3)		73	(4)
Abdominal tenderness: Yes	1205	787	(65)	1701	1228	(72)
Liver size [cm]	1195	2	(1-2)	1683	2	(1-2)
HCT [%]	1186	50	(47-52)	1683	49	(46-52)
PLT [1000 cell/mm3]	1188	38	(26-54)	1683	41	(28-60)
AST [IU/l]	910	133	(89-218)	1021	125	(80-206)
RT-PCR performed	1167			1635		
- DENV-1		658	(56)		672	(41)
- DENV-2		281	(24)		363	(22)
- DENV-3		19	(2)		46	(3)
- DENV-4		7	(1)		109	(7)
- Mixed		8	(1)		9	(1)
- Negative		194	(17)		436	(27)
Outcomes						
Used colloid: Yes	1207	544	(45)	1706	958	(56)
At least 2 colloid boluses	1207	218	(18)	1706	355	(21)
Total colloid volume [ml/kg]	544	19	(12-25)	958	25	(15-29)
Survival status: Died	1207	7	(1)	1706	8	(<1)
Major complications: Yes	1207	17	(1)	1706	28	(2)
Inotropic drug: Yes	1207	54	(4)	1706	74	(4)
Recurrent shock: Yes	1207	433	(36)	1706	593	(35)
Profound DSS: Yes	1207	222	(18)	1706	364	(21)
Critical DSS: Yes	1207	57	(5)	1706	83	(5)

Summary statistics are median (IQR) for continuous variables and frequency (%) for categorical variables.
Abbreviations: AST = aspartate aminotransferase, RT-PCR = reverse transcriptase polymerase chain reaction, DENV = dengue virus, DSS = dengue shock syndrome.

4.3.2 Analysis of profound DSS

Univariate analysis

Apart from bleeding, abdominal tenderness, liver size and platelet count, all other parameters assessed showed significant associations with profound DSS in the univariate analysis of the observational study population (Table 4.4). Results based on all 1706 patients were largely consistent.

Multivariable analysis

Linearity and additivity assessments As shown in Table 4.5, flexible spline functions showed a significant improvement for modelling the day of illness and HCT while linear terms seemed to be adequate for other covariates. Consistent with these findings, plots showing the estimated adjusted association of covariates with outcome from a generalized additive model (GAM) also indicated potential non-linearity in the relationships between day of illness and HCT with severity of disease (Figure 4.4). However, these non-linear associations were driven by rare patients with highly unusual covariate values. Indeed, for the day of illness at shock, the increase of severity from day 7 onwards just represents the high proportion of profound DSS amongst 4 unusual patients who had shock later than day 7 (2 out of 4 had profound DSS). Similarly, for HCT, the high proportion of profound DSS amongst 5 patients who had HCT less than 40% (3 out of 5 cases had profound DSS) may distort the relationship between HCT and severity. Of note, 2 out of these profound DSS cases had received intravenous fluid before enrolment into the study. The plots from a GAM-fit without these 9 unusual cases estimated a linear association with outcome, i.e. confirmed the adequacy of the simple model (Figure 4.5). Because the day of illness at shock and HCT values of the unusual cases appeared correct and plausible, they were not excluded from the analysis even though these cases are not typical for dengue shock patients. However, due to their low number, they do not provide convincing evidence for a non-linear association.

The assessment of pre-defined interaction terms revealed a significant interaction between haemodynamic index and gender (Table 4.5). Haemodynamic index is a categorical variable created by categorizing and combining systolic blood pressure and pulse pres-

sure. To exclude that the interaction is an artifact from categorization in the definition of haemodynamic index categories, alternative models which included systolic BP and pulse pressure as continuous covariates were fitted. This revealed that the interaction between haemodynamic index and gender can be explained by the interaction between systolic BP and gender and that the interaction remained when systolic BP was modelled as a continuous variable (p values of interaction tests are 0.05 and 0.02 for patients in the observational study including all subjects or only those with positive systolic BP only, respectively; of note, for the analysis of all subjects, systolic BP was modeled with 2 variables: continuous systolic BP and an indicator of zero systolic BP).

Based on these considerations, no non-linear terms were added to the multivariable logistic regression model but I added an interaction term between haemodynamic index and gender.

Table 4.4. *Univariate effects of candidate predictors on profound DSS estimated from univariate logistic regression models.*

Covariate	Observational study (n = 1207)			All patients (n = 1706)		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.85	(0.81, 0.89)	<0.01	0.83	(0.80, 0.86)	<0.01
Gender			0.04			0.07
- Female	1.00					1.00
- Male	0.74	(0.55, 0.99)		0.80	(0.64, 1.02)	
Weight [+5 kg]	0.84	(0.78, 0.90)	<0.01	0.83	(0.77, 0.88)	<0.01
Day of illness [+1 day]	0.73	(0.62, 0.87)	<0.01	0.72	(0.63, 0.82)	<0.01
Temperature [+1°C]	1.47	(1.07, 1.99)	0.02	1.49	(1.16, 1.90)	<0.01
Pulse rate [+10 per min]	1.14	(1.10, 1.19)	<0.01	1.11	(1.07, 1.14)	<0.01
Systolic BP [+5 mmHg]	0.92	(0.90, 0.94)	<0.01	0.93	(0.91, 0.95)	<0.01
Pulse pressure [+5 mmHg]	0.61	(0.55, 0.68)	<0.01	0.67	(0.61, 0.74)	<0.01
Haemodynamic index			<0.01			<0.01
- Group 1	1.00			1.00		
- Group 2	1.67	(1.19, 2.32)		1.62	(1.24, 2.11)	
- Group 3	5.06	(3.11, 8.23)		3.76	(2.53, 5.57)	
Haemorrhage			0.43			0.02
- None	1.00			1.00		
- Skin only	0.82	(0.60, 1.12)		0.80	(0.62, 1.05)	
- Mucosal	0.98	(0.41, 2.11)		1.57	(0.90, 2.68)	
Abdominal tenderness			0.29			0.21
- Yes	1.00			1.00		
- No	1.18	(0.87, 1.59)		1.19	(0.91, 1.56)	
Liver size [+1 cm]	1.04	(0.89, 1.21)	0.65	1.03	(0.91, 1.17)	0.62
HCT [+1 %]	1.07	(1.03, 1.11)	<0.01	1.06	(1.02, 1.09)	<0.01
PLT [+10,000 cell/mm^3]	1.03	(1.00, 1.07)	0.07	1.02	(0.99, 1.05)	0.19
PLT [per 10-fold increase]	1.58	(0.91, 2.76)	0.11	1.42	(0.91, 2.22)	0.13
AST [+1 IU/l]	1.00	(1.00, 1.00)	<0.01	1.00	(1.00, 1.00)	<0.01
AST [per 2 times increase]	1.25	(1.07, 1.45)	<0.01	1.27	(1.10, 1.47)	<0.01
RT-PCR result			0.23			0.04
- DENV-1	1.00			1.00		
- DENV-2	1.15	(0.81, 1.61)		1.22	(0.88, 1.69)	
- DENV-3	2.46	(0.90, 6.25)		2.92	(1.48, 5.71)	
- DENV-4	3.17	(0.62, 14.54)		1.28	(0.68, 2.39)	
- Mixed	0.60	(0.03, 3.43)		0.56	(0.03, 3.08)	

The analyses in all patients were adjusted for the randomized treatment assignment. Patients with negative RT-PCT were excluded in the analysis for RT-PCR.

Abbreviations: OR = odds ratio, CI = confidence interval, BP = blood pressure, HCT = haematocrit, PLT = platelet count, AST = aspartate aminotransferase, RT-PCR = reverse transcriptase polymerase chain reaction, DENV = dengue virus.

Table 4.5. *Linearity and additivity tests in the pre-defined logistic regression model for profound DSS (n = 1207).*

Predictor	Deviance	df	p value
Linearity tests (compared to a quadratic function)			
Age	1.40	1	0.24
Weight	<0.01	1	0.98
Day of illness	2.51	1	0.11
Pulse rate	0.44	1	0.51
Temperature	0.57	1	0.45
HCT	4.35	1	0.04
PLT	0.13	1	0.72
Linearity tests (compared to a natural cubic spline with 4 degrees of freedom)			
Age	6.83	3	0.08
Weight	4.36	3	0.22
Day of illness	8.46	3	0.04
Pulse rate	1.29	3	0.73
Temperature	2.83	3	0.42
HCT	7.99	3	0.05
PLT	0.41	3	0.94
Additivity tests (interaction tests)			
Age and all other covariates	16.7	14	0.27
Day of illness at shock and all other covariates	22.2	14	0.07
Haemodynamic index and all other covariates	35.6	24	0.06
Haemodynamic index and sex	9.9	2	0.01
Systolic BP and gender[a]	6.1	2	0.05
Systolic BP and gender[b]	5.7	1	0.02

[a] *Systolic BP includes continuous systolic BP and indicator of zero systolic BP, model on all patients in the observational study.*

[b] *Systolic BP just includes continuous systolic BP, model on patients in the observational study with positive systolic BP.*

Abbreviations: HCT = haematocrit, PLT = platelet count, BP = blood pressure.

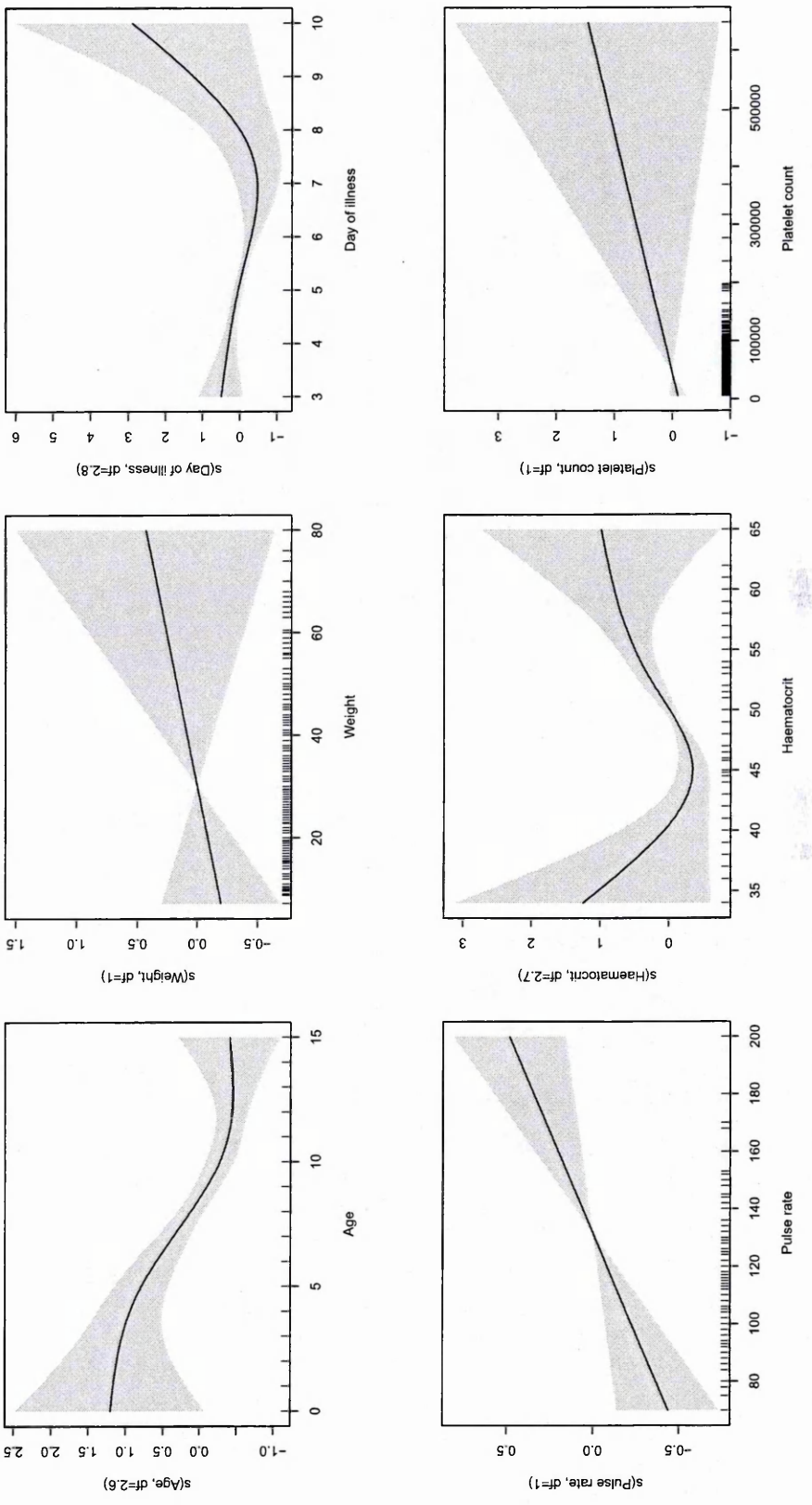

Figure 4.4. *Plots of estimated component smooth functions of a generalized addictive model (GAM) fit for profound DSS with continuous covariates modelled using natural cubic spline functions and integrated smoothness estimation. Dots correspond to individual partial residuals; solid lines correspond to spline functions estimated by GAM; gray areas correspond to point-wise 95% confidence intervals of the estimated values.*

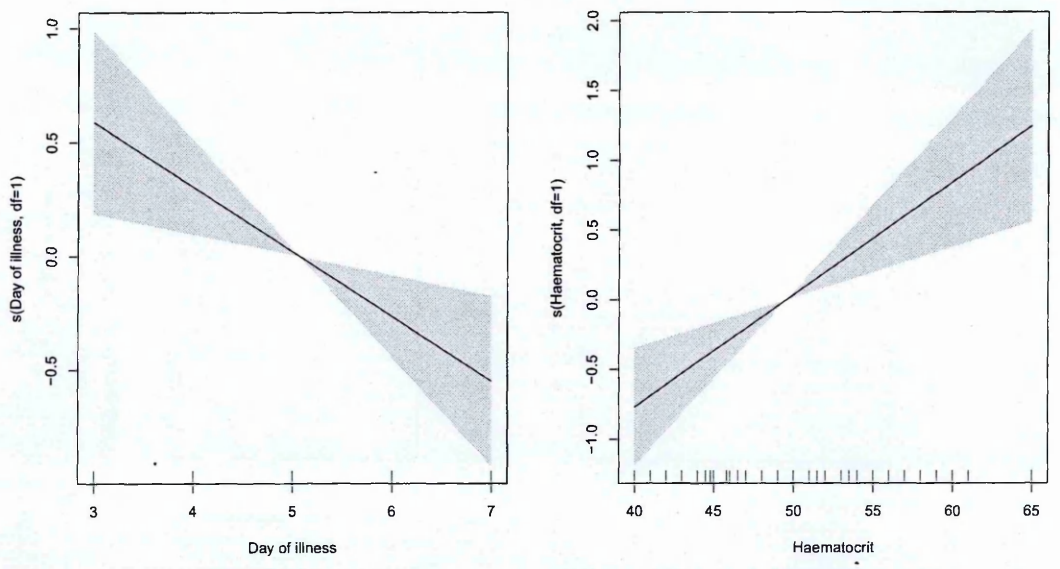

Figure 4.5. *Plots of estimated component smooth functions for day of illness and haematocrit from a generalized addictive model (GAM) fit for profound DSS after removal of 9 patients with day of illness > 7 or haematocrit values < 40%. Dots correspond to individual partial residuals; solid lines correspond to spline functions estimated by GAM; gray areas correspond to point-wise 95% confidence intervals of the estimated values.*

The full and reduced logistic regression model The multivariable models assessing relationships for the predefined candidate predictors with the primary outcome on the primary analysis population and on all patients are summarized in Table 4.6 & 4.7. Younger age, earlier day of illness at shock, faster pulse rate, higher temperature, higher haematocrit and worse haemodynamic status in females were all associated with profound DSS. These predictors were also chosen by the reduced model based on stepwise variable selection or best subset selection using AIC which both selected the same variables (Table 4.6 & 4.7).

Table 4.6. *Adjusted effects of candidate predictors on profound DSS estimated from the full logistic regression model and the reduced model with stepwise variable selection based on AIC (n = 1207).*

Covariate	Full model			Reduced model		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.86	(0.80, 0.93)	<0.01	0.87	(0.83, 0.92)	<0.01
Weight [+5 kg]	1.03	(0.92, 1.14)	0.59	-	-	-
Day of illness [+1 day]	0.79	(0.65, 0.94)	0.01	0.78	(0.65, 0.94)	<0.01
Temperature [+1°C]	1.58	(1.12, 2.21)	0.01	1.59	(1.12, 2.20)	<0.01
Pulse rate [+10 per min]	1.08	(1.03, 1.13)	<0.01	1.07	(1.03, 1.13)	<0.01
Haemorrhage			0.81			-
- None	1.00			-		
- Skin only	0.94	(0.67, 1.33)		-	-	
- Mucosal	1.22	(0.47, 2.89)		-	-	
Abdominal tenderness			0.91			-
- Yes	1.00			-		
- No	1.02	(0.70, 1.47)		-	-	
Liver size [+1 cm]	0.96	(0.80, 1.16)	0.69	-	-	-
HCT [+1 %]	1.07	(1.03, 1.12)	<0.01	1.07	(1.03, 1.12)	<0.01
PLT [+10,000 cell/mm^3]	1.02	(0.98, 1.06)	0.24	-	-	-
Gender						
- Female	1.00			1.00		
- Male	1.14	(0.75, 1.73)		1.17	(0.78, 1.77)	
Haemodynamic index – females						
- Group 1	1.00			1.00		
- Group 2 vs. group 1	2.57	(1.59, 4.15)		2.55	(1.58, 4.12)	
- Group 3 vs. group 1	3.01	(1.43, 6.36)		3.06	(1.46, 6.45)	
Haemodynamic index – males						
- Group 1	1.00			1.00		
- Group 2 vs. group 1	0.82	(0.46, 1.41)		0.79	(0.45, 1.36)	
- Group 3 vs. group 1	1.60	(0.70, 3.61)		1.55	(0.67, 3.49)	

95% confidence intervals for the reduced model do not take into account the uncertainty of model selection.
p values for gender and haemodynamic index are not provided due to interaction.
Abbreviations: OR = odds ratio, CI = confidence interval, HCT = haematocrit, PLT = platelet count, AIC = Akaike information criterion.

Table 4.7. *Adjusted effects of candidate predictors on profound DSS estimated from the full logistic regression model and the reduced model with stepwise variable selection based on AIC for all patients (n = 1706).*

Candidate predictors	Full model			Reduced model		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.82	(0.77, 0.87)	<0.01	0.84	(0.80, 0.88)	<0.01
Weight [+5 kg]	1.05	(0.96, 1.16)	0.27	-	-	-
Day of illness [+1 day]	0.74	(0.64, 0.85)	<0.01	0.74	(0.65, 0.86)	<0.01
Temperature [+1°C]	1.50	(1.14, 1.96)	<0.01	1.50	(1.14, 1.95)	<0.01
Pulse rate [+10 per min]	1.04	(1.00, 1.08)	0.08	1.04	(1.00, 1.08)	0.08
Haemorrhage			0.02			0.02
- None	1.00			1.00		
- Skin only	0.96	(0.72, 1.28)		0.94	(0.71, 1.25)	
- Mucosal	2.16	(1.17, 3.91)		2.19	(1.20, 3.95)	
Abdominal tenderness			0.47			-
- Yes	1.00			-		
- No	0.89	(0.65, 1.23)		-	-	
Liver size [+1 cm]	0.98	(0.85, 1.13)	0.79	-	-	-
HCT [+1 %]	1.07	(1.03, 1.11)	<0.01	1.07	(1.03, 1.11)	<0.01
PLT [+10,000 cell/mm^3]	1.01	(0.98, 1.04)	0.53			
Gender						
- Female	1.00			1.00		
- Male	1.13	(0.81, 1.58)		1.16	(0.84, 1.62)	
Haemodynamic index – females						
- Group 1	1.00			1.00		
- Group 2 vs. group 1	2.13	(1.44, 3.14)		2.11	(1.43, 3.10)	
- Group 3 vs. group 1	2.94	(1.59, 5.44)		2.93	(1.59, 5.43)	
Haemodynamic index – males						
- Group 1	1.00			1.00		
- Group 2 vs. group 1	1.06	(0.69, 1.61)		1.02	(0.66, 1.53)	
- Group 3 vs. group 1	1.15	(0.56, 2.31)		1.12	(0.55, 2.26)	

These analyses were adjusted for the randomized treatment assignment. 95% confidence intervals for the reduced model do not take into account the uncertainty of model selection.

p values for gender and haemodynamic index are not provided due to interaction.

Abbreviations: OR = odds ratio, CI = confidence interval, HCT = haematocrit, PLT = platelet count, AIC = Akaike information criterion.

Model performance As described in Table 4.8, all approaches have similar overall performance, in terms of the Brier score. However, their Brier scores are close to Brier scores of non-informative models that assign the incidence of the outcome as the predicted risk for all patients (estimated Brier score of non-informative models in temporal validation is 0.12; median [IQR] Brier score of non-informative models in internal validation is 0.15 [0.14-0.16]). Of note, Brier scores in temporal validation are lower than in internal validation, which may be explained by the lower outcome incidence in the test set in temporal validation (incidence of profound DSS in test set is 0.15 in temporal validation; in internal validation, median [IQR] outcome incidence is 0.18 [0.16-0.21]). The reduced models had very similar performance characteristics to the full model in terms of discrimination and calibration. Both models showed acceptable discrimination for both temporal and internal validation with an AUC of at least 0.69. Calibration in internal validation was also satisfactory. However, in temporal validation, models developed using data from the 939 patients enrolled before 2009 tended to overestimate the risk of profound DSS for the 268 patients enrolled in 2009 (i.e. the observed incidence of this outcome in the test set was 15% compared to an average predicted risk of 21% in both the full and reduced models) (Table 4.8 & Figure 4.6). The reduced model performed better than, or as well as, alternative logistic models and more sophisticated models including lasso, GAM, CART and boosting (Table 4.8). While several alternative models showed satisfactory performance, CART showed quite poor results.

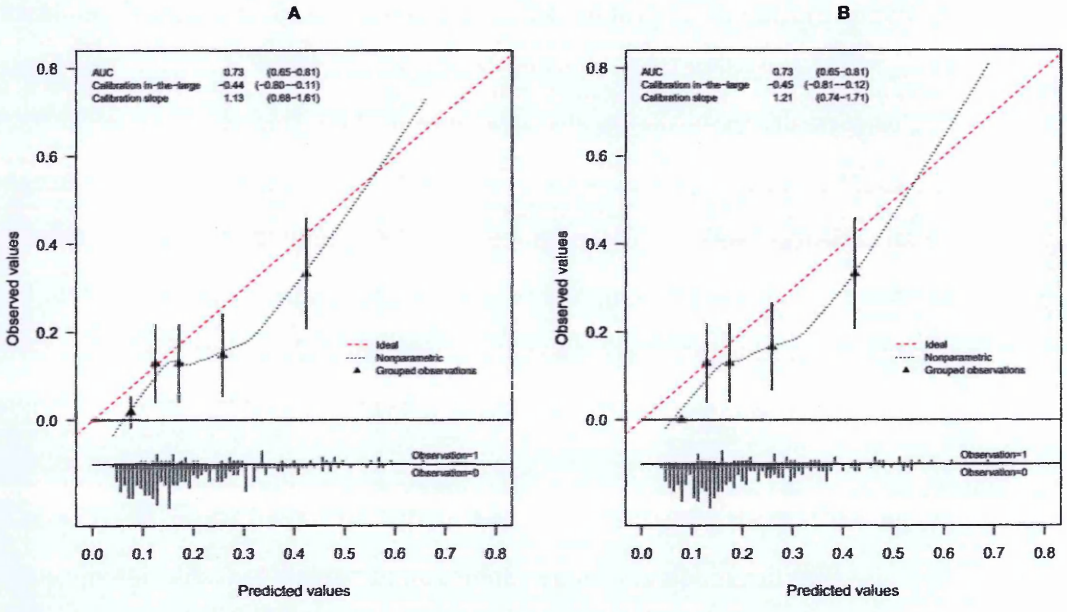

Figure 4.6. *Calibration plots for temporal validation of the full logistic model (panel A) and the reduced model with variable selection based on AIC (panel B) for profound DSS. Model development was on patients enrolled before 2009 (n = 939), validation on patients enrolled in 2009 (n = 268). Black triangles show average predicted versus observed risk for 5 patient strata of equal size grouped according to their predicted risk. Corresponding vertical lines show 95% confidence intervals. The black line corresponds to a non-parametric smoother of the predicted versus observed values. The red dashed reference line corresponds to the ideal relationship. Each panel also describes temporal validated performance of each model (left upper corner) and the distribution of observed values (at the bottom).*

Table 4.8. Performance of alternative models for profound DSS (n = 1207).

Measure	Logistic 1	Logistic 2	Logistic 3	Logistic 4	Lasso	CART	GAM	Boosting
Temporal validation								
Brier score	0.12	0.12	0.11	0.12	0.12	0.15	0.12	0.12
AUC	0.71	0.73	0.74	0.68	0.71	0.61	0.69	0.72
(95% CI)	(0.62, 0.79)	(0.64, 0.81)	(0.65, 0.82)	(0.59, 0.78)	(0.63, 0.80)	(0.52, 0.71)	(0.61, 0.78)	(0.63, 0.80)
Calibration-in-the-large	-0.46	-0.49	-0.50	-0.54	-0.46	-0.60	-0.47	-0.44
(95% CI)	(-0.83, -0.12)	(-0.86, -0.15)	(-0.87, -0.16)	(-0.90, -0.20)	(-0.83, -0.12)	(-0.99, -0.23)	(-0.84, -0.12)	(-0.80, -0.10)
Calibration slope	1.06	1.13	1.22	0.99	1.16	0.37	0.83	1.21
(95% CI)	(0.61, 1.54)	(0.68, 1.62)	(0.76, 1.73)	(0.52, 1.49)	(0.67, 1.67)	(0.05, 0.67)	(0.44, 1.26)	(0.68, 1.76)
Internal validation								
Brier score	0.14	0.14	0.14	0.14	0.14	0.15	0.14	0.14
AUC	0.68	0.69	0.69	0.67	0.67	0.61	0.68	0.70
Calibration-in-the-large	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.03
Calibration slope	0.89	0.87	0.92	0.89	1.03	0.36	0.79	1.17

Logistic 1: full logistic model without interaction between haemodynamic index and gender; Logistic 2: full logistic model with interaction (full model in the text); Logistic 3: logistic model with interaction and subsequent stepwise backwards model selection based on AIC (reduced model in the text); Logistic 4: logistic model with interaction and subsequent stepwise backwards model selection based on BIC.

Abbreviations: AUC = area under the ROC curve, CI = confidence interval, CART = classification and regression trees, GAM = generalized additive models.

The score chart

A simplified score chart for clinical use based on the reduced model was developed following the procedures described in section Model presentation, and is shown in Figure 4.7. The base model was the reduced logistic model (stepwise model selection using AIC) displayed in Table 4.6. However, results in Table 4.6 give lower risk predictions in males with haemodynamic index group 2 compared to group 1. As the corresponding difference in estimates is small, non-significant, and clinically implausible, haemodynamic index groups 1 and 2 in males were pooled, and then the logistic model was refitted prior to deriving the point score. Of note, this score chart was derived without any adjustment for mis-calibration. However, this adjustment could have been done by applying the internally-validated calibration intercept and slope to the linear predictors before deriving the final score.

This score chart assigns points to each predictor value and the total point sum is then translated to the predicted risk of a severe outcome. For example, the total point sum for a 10-year-old girl who presents on day 6 of illness with a pulse rate of 100 beats/min, a temperature of 37.5°C, a haematocrit of 44% and a haemodynamic index of 1 is 11 (5+2+1+2+1+0), and therefore, her estimated risk of developing profound shock during hospitalization is less than 10%. Based on this low estimated risk and taking into account the available resources and clinical expertise, the treating physician may decide to keep this patient in his/her health facility with an appropriate monitoring schedule rather than to refer the patient to a higher level hospital.

The adequacy of this score chart was evaluated by comparing risk predictions from the score chart to those from the logistic model with AIC model selection for patients in the observational study. The median (IQR) of the differences between these two risk-estimation approaches was 0.018 (0.004, 0.035) and the range was -0.093 to 0.126. The largest differences occurred in patients with intermediate predicted risks (Figure 4.8 & 4.9).

Point score

Risk factor	Categories	Points	Risk factor	Categories	Points
Age [years]		15 − age	Temperature [°C]	< 37.5	0
				37.5 − 37.9	2
				38.0 − 38.4	3
				38.5 − 38.9	5
				≥ 39	6
Day of illness	< 4	7	Haematocrit [%]	< 40	0
	4	5		40 − 44	1
	5	4		45 − 49	4
	6	2		50 − 54	7
	≥ 7	0		55 − 59	9
				≥ 60	10
Pulse [per min]	< 90	0	Hemodynamic Index		
	90 − 109	1	For females	1	0
	110 − 129	2		2	7
	130 − 149	3		3	8
	150 − 169	4			
	170 − 189	5	For males	1 or 2	1
	Rapid, thready	6		3	4

Risk estimation

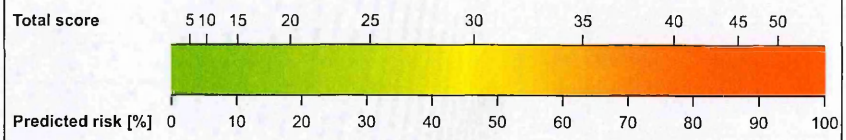

Figure 4.7. *Score-chart for prediction of profound DSS. The upper panel assigns a point score for each risk factor while the lower panel assigns the predicted risk of developing profound shock based on the total point sum for all risk factors.*

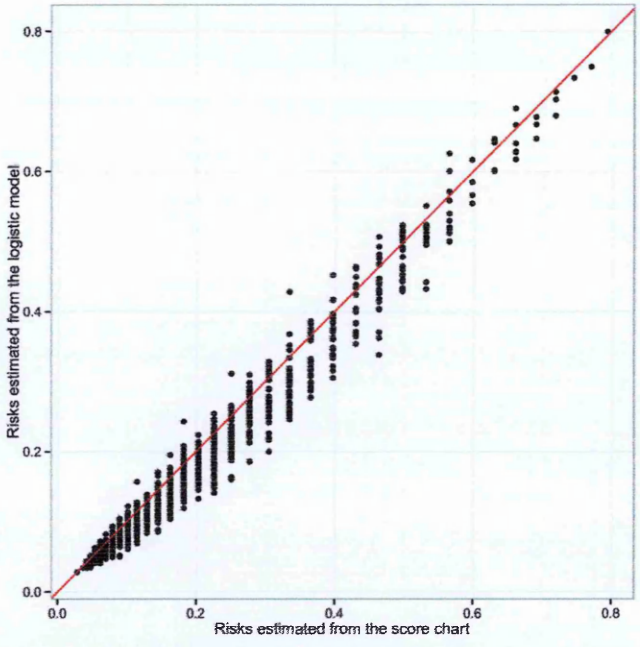

Figure 4.8. *Scatter plot of risks estimated from the score chart versus the logistic regression model.*

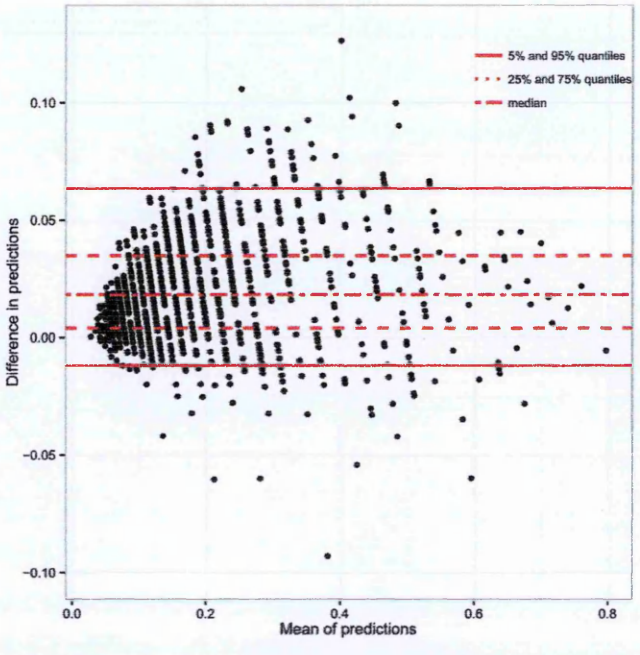

Figure 4.9. *Bland-Altman plot of differences between risks estimated by the score chart and those estimated by the logistic model for profound DSS with Akaike information criterion (AIC) model selection (and without pooling the haemodynamic index categories in males).*

4.3.3 Analysis of recurrent shock

Results from the analysis of recurrent shock were consistent with results from the analysis of profound DSS and for both outcomes; there was a significant interaction between haemodynamic index and gender (Table 4.9). Independent risk factors of recurrent shock were younger age, earlier day of illness at shock, higher temperature, faster pulse rate, higher haematocrit and worse haemodynamic status in females. The prediction model for recurrent shock based on these risk factors had a moderate performance and AUC, calibration-in-the-large, and calibration slope in internal validation were 0.64, -0.003, 0.86 (for the full model) and 0.65, -0.004, 0.93 (for the reduced model with stepwise variable selection based on AIC), respectively (Table 4.10).

This dataset also provides an opportunity to externally assess the performance of a prognostic model developed by Huy et al. (2013a). Their model aimed to predict the occurrence of recurrent shock based on five variables: day of illness at hospital admission, presence of purpura/ecchymosis and ascites/pleural effusion at shock, as well as platelet count and pulse pressure at shock. For the DF data, the model equation of Huy et al. (2013a) yielded an AUC of 0.54 (95% confidence interval: 0.50-0.57) and substantially under-estimated the true risk of recurrent shock (average predicted risk 15% compared to an observed risk of 36%).

Table 4.9. *Adjusted effects of candidate predictors for recurrent shock estimated from logistic models for patients in the observational study and all patients.*

Covariate	Observational study			All patients		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.90	(0.85, 0.96)	<0.01	0.86	(0.82, 0.91)	<0.01
Weight [+5 kg]	1.00	(0.92, 1.08)	0.96	1.03	(0.96, 1.11)	0.46
Day of illness [+1 day]	0.78	(0.68, 0.91)	<0.01	0.75	(0.66, 0.84)	<0.01
Temperature [+1°C]	1.85	(1.39, 2.48)	<0.01	1.79	(1.41, 2.28)	<0.01
Pulse rate [+10 per min]	1.07	(1.02, 1.11)	<0.01	1.05	(1.01, 1.08)	0.01
Haemorrhage			0.95			0.19
- None	1.00			1.00		
- Skin only	0.96	(0.73, 1.26)		0.97	(0.76, 1.23)	
- Mucosal	1.02	(0.47, 2.14)		1.57	(0.90, 2.71)	
Abdominal tenderness			0.76			0.98
- Yes	1.00			1.00		
- No	0.95	(0.71, 1.28)		1.00	(0.77, 1.31)	
Liver size [+1 cm]	1.05	(0.91, 1.22)	0.49	1.03	(0.91, 1.17)	0.63
HCT [+1 %]	1.07	(1.03, 1.11)	<0.01	1.07	(1.04, 1.10)	<0.01
PLT [+10,000 cell/mm3]	1.00	(0.96, 1.03)	0.86	0.99	(0.97, 1.02)	0.69
Gender						
- Female	1.00			1.00		
- Male	1.42	(1.04, 1.95)		1.40	(1.07, 1.83)	
Haemodynamic index – females						
- Group 1	1.00			1.00		
- Group 2 vs. group 1	1.86	(1.24, 2.79)		1.77	(1.26, 2.50)	
- Group 3 vs. group 1	1.51	(0.75, 3.05)		1.74	(0.98, 3.11)	
Haemodynamic index – males						
- Group 1	1.00			1.00		
- Group 2 vs. group 1	0.77	(0.50, 1.16)		0.91	(0.64, 1.28)	
- Group 3 vs. group 1	0.65	(0.30, 1.41)		0.56	(0.29, 1.08)	

Interaction tests between gender and haemodynamic index were significant in both models (p<0.01 in both models). p values for gender and haemodynamic index are not provided due to interaction.

The analyses in all patients were adjusted for the randomized treatment assignment.

Abbreviations: OR = odds ratio, CI = confidence interval, HCT = haematocrit, PLT = platelet count.

Table 4.10. Performance of alternative models for recurrent shock (n = 1207).

Measure	Logistic 1	Logistic 2	Logistic 3	Logistic 4	Lasso	CART	GAM	Boosting
Temporal validation								
Brier score	0.18	0.18	0.18	0.18	0.18	0.21	0.19	0.18
AUC	0.67	0.69	0.71	0.69	0.68	0.61	0.65	0.67
(95% CI)	(0.58, 0.76)	(0.60, 0.78)	(0.62, 0.79)	(0.60, 0.78)	(0.59, 0.77)	(0.52, 0.70)	(0.56, 0.74)	(0.58, 0.77)
Calibration-in-the-large	-1.45	-1.47	-1.47	-1.46	-1.41	-1.57	-1.47	-1.39
(95% CI)	(-1.81, -1.11)	(-1.83, -1.14)	(-1.83, -1.13)	(-1.82, -1.12)	(-1.77, -1.08)	(-1.95, -1.22)	(-1.83, -1.13)	(-1.75, -1.05)
Calibration slope	1.09	1.20	1.34	1.26	1.41	0.27	0.77	1.26
(95% CI)	(0.52, 1.69)	(0.63, 1.82)	(0.75, 1.97)	(0.66, 1.91)	(0.74, 2.14)	(-0.11, 0.62)	(0.27, 1.30)	(0.58, 1.97)
Internal validation								
Brier score	0.22	0.22	0.21	0.22	0.22	0.23	0.22	0.22
AUC	0.64	0.64	0.65	0.64	0.65	0.57	0.64	0.65
Calibration-in-the-large	-0.004	-0.003	-0.004	-0.006	-0.005	0.002	-0.006	-0.004
Calibration slope	0.88	0.86	0.93	0.96	1.16	0.49	0.82	1.19

Logistic 1: full logistic model without interaction between haemodynamic index and gender; Logistic 2: full logistic model with interaction ; Logistic 3: logistic model with interaction and subsequent stepwise backwards model selection based on AIC; Logistic 4: logistic model with interaction and subsequent stepwise backwards model selection based on BIC.

Abbreviations: AUC = area under the ROC curve, CI = confidence interval, CART = classification and regression trees, GAM = generalized additive models.

4.3.4 Analysis of critical DSS

The numbers of patients experiencing critical DSS in the observational study and in the whole patient population were 57/1207 (5%) and 83/1706 (5%), respectively. The multivariable analysis of this outcome in the observational study patients identified similar risk factors as for the profound DSS (Table 4.11) except that pulse rate was not an independent risk factor, and there was no evidence of an interaction between gender and haemodynamic index.

Table 4.11. *Adjusted effects of candidate predictors on critical DSS estimated from logistic models for patients in the observational study and for all patients.*

Covariate	Observational study			All patients		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.73	(0.62, 0.86)	<0.01	0.72	(0.62, 0.82)	<0.01
Gender			0.09			0.03
- Female	1.00			1.00		
- Male	0.59	(0.32, 1.08)		0.59	(0.36, 0.96)	
Weight [+5 kg]	0.96	(0.73, 1.22)	0.73	1.00	(0.80, 1.24)	0.97
Day of illness [+1 day]	0.71	(0.49, 1.00)	0.05	0.77	(0.59, 1.01)	0.06
Temperature [+1°C]	1.51	(0.84, 2.54)	0.16	1.83	(1.20, 2.72)	<0.01
Pulse rate [+10 per min]	0.97	(0.87, 1.06)	0.48	0.98	(0.90, 1.06)	0.62
Haemodynamic index			0.04			0.23
- Group 1	1.00			1.00		
- Group 2	1.00	(0.47, 1.99)		1.04	(0.59, 1.81)	
- Group 3	3.36	(1.26, 9.18)		2.13	(0.89, 5.12)	
Haemorrhage			0.63			0.17
- None	1.00			1.00		
- Skin only	0.75	(0.41, 1.39)		0.89	(0.53, 1.51)	
- Mucosal	1.06	(0.14, 5.01)		2.42	(0.80, 6.52)	
Abdominal tenderness			0.78			0.44
- Yes	1.00			1.00		
- No	1.10	(0.56, 2.10)		1.26	(0.70, 2.23)	
Liver size [+1 cm]	1.00	(0.71, 1.41)	0.99	1.13	(0.86, 1.49)	0.38
HCT [+1 %]	1.10	(1.02, 1.19)	0.01	1.09	(1.02, 1.16)	<0.01
PLT [+10,000 cell/mm^3]	1.04	(0.99, 1.10)	0.10	1.04	(1.00, 1.09)	0.06

Interaction tests between gender and haemodynamic index were not significant for both models (p=0.45 and 0.85, respectively). The analyses in all patients were adjusted for the randomized treatment assignment.

Abbreviations: OR = odds ratio, CI = confidence interval, HCT = haematocrit, PLT = platelet count.

4.3.5 Analysis of total volume of colloid

Results from Cox regression identified the same risk factors for the higher volume of colloid required (Table 4.12 & 4.13). These results were also in agreement with the analysis of primary endpoint, at least in the direction of potential effects of pre-defined covariates on outcome.

Table 4.12. *Adjusted effects of candidate predictors on the total volume of colloid estimated from Cox models for patients in observational study (n = 1207).*

Covariate	Model 1			Model 2		
	HR	95% CI	p value	HR	95% CI	p value
Age [+1 year]	1.06	(1.03, 1.09)	<0.01	1.06	(1.03, 1.09)	<0.01
Gender			0.78			0.75
- Female	1.00			1.00		
- Male	1.02	(0.90, 1.15)		1.02	(0.90, 1.15)	
Weight [+5 kg]	1.04	(1.00, 1.07)	0.07	1.03	(0.99, 1.07)	0.09
Day of illness [+1 day]	1.11	(1.04, 1.18)	<0.01	1.11	(1.04, 1.18)	<0.01
Temperature [+1°C]	0.85	(0.74, 0.98)	0.02	0.82	(0.71, 0.94)	<0.01
Pulse rate [+10 per min]	0.96	(0.94, 0.98)	<0.01	0.96	(0.94, 0.98)	<0.01
Haemodynamic index			<0.01			<0.01
- Group 1	1.00			1.00		
- Group 2	0.90	(0.78, 1.03)		0.89	(0.78, 1.02)	
- Group 3	0.55	(0.42, 0.71)		0.57	(0.44, 0.74)	
Haemorrhage			0.67			0.62
- None	1.00			1.00		
- Skin only	1.06	(0.93, 1.20)		1.06	(0.93, 1.20)	
- Mucosal	1.06	(0.76, 1.48)		0.98	(0.71, 1.37)	
Abdominal tenderness			0.24			0.18
- Yes	1.00			1.00		
- No	1.08	(0.95, 1.24)		1.10	(0.96, 1.25)	
Liver size [+1 cm]	1.02	(0.96, 1.09)	0.55	1.01	(0.95, 1.08)	0.68
HCT [+1 %]	0.97	(0.96, 0.98)	<0.01	0.97	(0.96, 0.98)	<0.01
PLT [+10,000 cell/mm^3]	1.00	(0.98, 1.01)	0.58	1.00	(0.98, 1.01)	0.72

Outcome of who died were either treated as right-censored observations (model 1) or replaced by the maximum observed outcome plus 1 (model 2).

Abbreviations: HR = hazards ratio, CI = confidence interval, HCT = haematocrit, PLT = platelet count.

Table 4.13. *Adjusted effects of candidate predictors on the total volume of colloid estimated from Cox regression models for all patients (n = 1706).*

Covariate	Model 1			Model 2		
	HR	95% CI	p value	HR	95% CI	p value
Age [+1 year]	1.07	(1.05, 1.10)	<0.01	1.07	(1.05, 1.10)	<0.01
Gender			0.40			0.36
- Female	1.00			1.00		
- Male	1.04	(0.94, 1.15)		1.05	(0.95, 1.16)	
Weight [+5 kg]	1.05	(1.01, 1.08)	0.01	1.04	(1.00, 1.08)	0.03
Day of illness [+1 day]	1.13	(1.07, 1.18)	<0.01	1.12	(1.07, 1.18)	<0.01
Temperature [+1°C]	0.82	(0.73, 0.92)	<0.01	0.78	(0.70, 0.88)	<0.01
Pulse rate [+10 per min]	0.97	(0.95, 0.98)	<0.01	0.96	(0.95, 0.98)	<0.01
Haemodynamic index			<0.01			<0.01
- Group 1	1.00			1.00		
- Group 2	0.92	(0.82, 1.03)		0.92	(0.82, 1.03)	
- Group 3	0.63	(0.50, 0.78)		0.66	(0.53, 0.82)	
Haemorrhage			0.84			0.69
- None	1.00			1.00		
- Skin only	1.03	(0.92, 1.15)		1.03	(0.92, 1.15)	
- Mucosal	1.00	(0.78, 1.29)		0.94	(0.73, 1.21)	
Abdominal tenderness			0.69			0.53
- Yes	1.00			1.00		
- No	1.03	(0.91, 1.16)		1.04	(0.92, 1.18)	
Liver size [+1 cm]	1.01	(0.96, 1.07)	0.72	1.00	(0.95, 1.06)	0.95
HCT [+1 %]	0.97	(0.96, 0.98)	<0.01	0.97	(0.96, 0.98)	<0.01
PLT [+10,000 cell/mm^3]	0.99	(0.97, 1.00)	0.05	0.99	(0.98, 1.00)	0.09

Outcome of who died were either treated as right-censored observations (model 1) or replaced by the maximum observed outcome plus 1 (model 2). These analyses were adjusted for the randomized treatment assignment.

Abbreviations: HR = hazards ratio, CI = confidence interval, HCT = haematocrit, PLT = platelet count.

4.4 Discussion

This chapter evaluated risk factors for poor outcomes in children with DSS. Younger age, earlier day of illness, higher temperature, faster pulse rate, higher haematocrit, and a worse haemodynamic status (in females) at onset of shock were associated with a higher risk of developing profound DSS, the primary outcome of this study. The results for secondary outcomes including recurrent shock, critical DSS and total volume of colloid, were largely consistent with the primary analysis. A robust prediction model for profound shock was developed and presented as a simple score-chart designed to assist decision-making in clinical practice.

The pathognomonic feature of the vasculopathy associated with severe dengue is an increase in intrinsic vascular permeability resulting in a transient capillary leak syndrome. Cardiovascular decompensation occurs when plasma losses exceed the capacity for up-regulation of the normal compensatory mechanisms that maintain plasma volume within well-circumscribed limits (Trung and Wills, 2010). Several studies have demonstrated a greater risk for vascular leakage and development of DSS among children compared to adults (Anders et al., 2011; Guzman et al., 2002; Hammond et al., 2005; Dinh The et al., 2012), probably related to higher intrinsic permeability with younger age (Gamble et al., 2000), and a relationship with severity of shock is therefore to be expected. Similarly earlier presentation with DSS implies more severe capillary leakage that quickly overwhelms the capacity for compensation, and this is consistent with the associations demonstrated between profound shock and other markers of leakage severity such as higher haematocrit and more severe haemodynamic compromise at presentation. Higher temperature at onset of shock was also associated with profound shock; 9% of cases had a temperature of 38 degrees or more at presentation irrespective of the day of illness (Chapter 3) possibly indicating a greater viral burden or a more intense immune response in these cases. Several of these factors, and others such as aspartate aminotransferase identified in the univariate analysis, have also been identified as risk factors for development of shock and/or more severe dengue disease generally (Anders et al., 2011; Potts et al., 2010a; Srikiatkhachorn and Green, 2010).

Interestingly no significant relationship between platelet count and shock severity was

found, although other studies have indicated a strong association with leakage severity (Wills et al., 2009). However, it is probable that the profound thrombocytopenia already present at enrolment masked any additional effects, and that a very large sample size would be required to detect differences between severity groups. Similarly the absence of any relationship between abdominal tenderness or liver size and shock severity likely reflects the fact that these parameters are closely linked with development of shock per se.

Female gender has been identified previously as an independent predictor of mortality in children with DSS, possibly reflecting higher intrinsic vascular permeability and thus greater susceptibility to capillary leak syndrome in females than males (Anders et al., 2011). In this study, there was a difference in the effect of initial haemodynamic status by gender in the analysis of profound DSS and recurrent shock but not with the more restricted definition of critical DSS. Potentially the higher intrinsic permeability in female subjects does influence the severity of leakage but only up to a critical point; when haemodynamic collapse finally occurs all subjects do badly and the differential effect of gender is obscured. Of note however, in the analysis of the secondary outcome of critical DSS, the event rate was lower and hence the statistical power to identify associations was reduced.

The final clinical prediction model provides a reliable tool to predict development of profound shock among DSS cases. The candidate predictors and primary outcome were prospectively defined, and the model was carefully developed and validated following standard methodology expected to minimize optimistic and/or spurious results (Harrell, 2001; Steyerberg, 2010). The final full model showed good calibration and discrimination, with an AUC of 0.74 (0.65, 0.82) for temporal validation, and performed favourably compared to a number of alternative modelling strategies. In temporal validation, all the statistical models tended to overestimate the average risk of profound DSS in patients recruited in 2009 by about 5%, compared to models developed on patients recruited earlier during the study-period. There was no systematic linear time-trends in the risk for profound DSS over time and the same recruitment protocol and treatment regime were used throughout the study. However, the observed risk in 2009 was the lowest over the entire observation period (Figure 4.3). Although the over-estimation in 2009 could be a chance finding it is also possible that some undefined change did occur, only becoming apparent

in 2009, but if so, the overall effect was minor.

The final model was simplified to a score chart. While this approach results in a mild loss of precision (Steyerberg, 2010), a chart is easier to understand than a regression formula or a nomogram, and allows clinicians to rapidly assess a patient's risk of progression. There is no clear-cut risk stratification or a decision rule based on the final prognostic model, since such a rule would require careful evaluation of costs and benefits, together with a defined intervention strategy for high-risk patients. Currently no such strategy exists, and effective management relies on careful monitoring and assiduous supportive care. In these circumstances experienced doctors are better equipped to decide on a particular therapeutic regimen, and the contribution of the current prognostic model is to provide physicians with guidance on the likely risk for developing profound shock. Ideally all DSS cases should be managed in a high-dependency unit (HDU) or ICU, but such facilities are limited especially in dengue-endemic areas, and given the very large numbers of potentially severe cases encountered daily, it can be difficult to prioritise individual cases. Using this score physicians may elect to monitor high-risk patients closely in a local HDU or ICU setting, or may choose to transfer them early to tertiary-level facilities, allowing more effective use of available staff and equipment for the remaining DSS cases. In the wider context, a prediction model such as this could be useful for identification of target populations for studies evaluating novel interventions for DSS (Simmons et al., 2012a).

Regarding prediction model for recurrent shock, the model developed in this project differs from the recent prediction model for recurrent shock developed by Huy et al. (2013a) in several respects, even though both models were based on Vietnamese children with DSS. Data of 1207 patients from a single hospital was used in this study whereas the study by Huy et al combined data of 444 patients from two very different health care settings (a preventive health-care centre of a small province and a large referral hospital in a big city) but did not report site-specific summaries. That model identified admission day, purpura/ecchymosis, ascites/pleural effusion, platelet count and pulse pressure, as risk factors for recurrent shock (Huy et al., 2013a); whereas identified risk factors for recurrent shock in this study were younger age, earlier day of illness at shock, higher temperature, faster pulse rate, higher haematocrit and worse haemodynamic status in females. It is difficult to compare the models, as the reported proportions of patients with

purpura/ecchymosis (36% in Huy's study vs. 3% in this study) and ascites/pleural effusion (44% in Huy's study vs. 1% in this study) were markedly different. Of note all data in this study were collected within 2 hours of onset of shock, while the timing of data collection in Huy's study was not clearly specified; the high incidence of features that typically develop after initiation of fluid resuscitation suggests that timing of data collection may be relevant, and may explain why application of Huy's model to this dataset showed only low discrimination (AUC 0.54) for the prediction of recurrent shock.

One potential limitation of this study relates to the definition of the clinical outcomes, some of which may be considered subjective. However, management of DSS in endemic areas is generally protocol driven, following a long-established precedent established by the World Health Organization (World Health Organization, 1997), and adherence to local management guidelines is typically very good in Vietnam. In addition, results from analysis of recurrent and profound shock were largely consistent with the analysis of critical DSS, which is less prone to clinician bias. Furthermore, in the context of DSS where prompt diagnosis with immediate fluid resuscitation is very effective (World Health Organization, 2009), the occurrence of hard outcomes such as death or major complications depends heavily on local expertise and the facilities of the healthcare system (Gibbons and Vaughn, 2002). Although the definition of profound DSS in this study might be a robust assessment of the overall severity of DSS, physicians applying these results must understand the provided reasoning as well as the potential pitfalls inherent in this type of analysis.

Developing a prognostic model using data from a single hospital with better expertise and facilities than many local healthcare facilities may be considered another limitation of the study, potentially restricting generalizability outside the primary context. However, as all the risk factors identified are clinically plausible, the model might also discriminate effectively in other settings. To adjust the model to provincial hospitals in Vietnam, where the distribution of predictors would likely be very similar to this study, simple re-calibration of the intercept of the current prognostic model to take into account differences in outcome prevalence may be sufficient, and such a re-calibration could be performed with a much smaller sample size (Steyerberg, 2010). However, re-calibration or re-estimation of the regression coefficients may be required to adapt the current model

to settings with markedly different patient characteristics, facilities or management guidelines (Steyerberg, 2010). Further work is needed to assess the performance of the model in a variety of hospitals and clinics within the region, as well as more broadly across healthcare systems in parts of the world where dengue infection is less common.

In summary, this chapter identified several clinical and laboratory risk factors of severe outcome amongst children with DSS. Based on these predictors, a simple score chart for profound DSS prediction was derived. This score-chart, which is simple to understand and easy to apply, could play a valuable role in triage and management of children with DSS in endemic areas, although precise prediction alone cannot improve clinical decision-making or overall outcomes.

Chapter 5

Prognostic models for DSS in hospitalized children with dengue

Summary

This chapter describes risk factors for DSS and presents a prognostic model for progression to DSS amongst hospitalized children with dengue using baseline information only.

5.1 Introduction

In contrast to the previous two chapters, which studied the DF cohort of children with dengue shock syndrome (DSS), this chapter examines the broader MD cohort of hospitalized children with dengue and investigates risk factors for progression to DSS. As described in Chapter 1, several earlier studies aimed to identify risk factors for severe outcomes amongst patients with dengue. However, only a few studies also tried to incorporate the identified risk factors into clinical prediction models in order to provide useful tools for clinical practice. Moreover, most of these studies looked at DHF rather than DSS as the primary outcome, included only small sample sizes, and used non-standardized modeling strategies.

In this study, we aimed to assess the predictive ability of several clinical and laboratory variables which are commonly available in endemic countries like Vietnam for the outcome of progression to DSS and to incorporate these variables into a prediction model using a large dataset of children hospitalized with dengue infection.

5.2 Methods

This chapter utilized data from the MD cohort. Detailed information related to study design, study participants, dengue diagnostics, general statistical analyses (descriptive analysis, treatment of missing values) and modeling strategy are described in Chapter 2. I present here definitions of the primary study population, clinical outcomes and candidate predictors and several specific statistical methods used in this chapter.

5.2.1 Study population

The primary study population included only patients with laboratory-confirmed dengue who were enrolled before day 5 of illness and did not experience DSS on the day of enrolment. Patients enrolled at a later day of illness were excluded as DSS most frequently occurs on day 5 or 6 of illness (Chapter 3). However, all patients regardless of the day of illness at enrolment were included in the descriptive analyses.

5.2.2 Clinical outcomes and candidate predictors

In this study, the primary outcome of interest is the occurrence of DSS. In addition, the following clinical outcomes during hospitalization were described: referral to the pediatric intensive care unit (PICU), new bleeding, requirement for intravenous fluids, the platelet nadir, the day of the platelet nadir, the maximum haematocrit (HCT), the day of the maximum HCT, and the overall haemoconcentration. Definitions of these clinical outcomes are described in Table 5.1.

Table 5.1. Definition of clinical outcomes.

Clinical outcome	Definitions
Dengue shock syndrome	WHO definition (World Health Organization, 2009)
Referral to the PICU	Being referred to the PICU
New bleeding	Having new bleeding during hospitalization
Platelet nadir	The minimum PLT count from day 3 to day 8 of illness (set to missing if <3 PLT values were available for a subject)
Day of platelet nadir	The day of illness at which the platelet nadir was reached
Maximum HCT	The maximum HCT value from day 3 to day 8 (set to missing if <3 HCT values were available for a subject)
Day of maximum HCT	The day of illness at which the maximum HCT was reached
Overall haemoconcentration	The overall haemoconcentration was defined as the percentage change of the maximum HCT from day 3 to 8 compared to the normal HCT for a specific patient. The normal HCT for a specific patient was defined as the HCT value at follow-up (after day 14 of illness). If this was not available, the minimum HCT value before day 2 of illness (provided the PLT count at the same time was $\geq$ 200,000 cells/mm^3) or the population value (37% for children from 5 to 10 years old, 38.5% for females more than 10 years old, 40% for males more than 10 years old) was used.

Abbreviations: PICU = paediatric intensive care unit, HCT = haematocrit, PLT = platelet count.

Table 5.2 describes candidate predictors measured at the time of enrolment into the cohort. These predictors included the presence of WHO warning signs (World Health Organization, 2009; Alexander et al., 2011) and other clinical signs and symptoms that were identified as risk factors of severe dengue in previous studies (Huy et al., 2013a). Serotype and immune status were only included in univariate analyses but not in the multivariable analysis as they were missing in a large number of participants and in general they would not be available to the treating physician in clinical practice.

Table 5.2. List of candidate predictors.

Predictor	Unit or possible values	Type
Age	Year	Continuous
Gender	Female/Male	Binary
Weight	Kg	Continuous
Day of illness	Day of illness at enrolment	Continuous
History of tiredness	Yes/No	Binary
History of vomiting	Yes/No	Binary
Tourniquet test	Negative/Equivocal/Positive	Categorical
Temperature	Body temperature [°C] measured in the axilla	Continuous
Pulse rate	Beats per minute	Continuous
Systolic blood pressure	mmHg	Continuous
Mucosal bleeding	Yes/No	Binary
Abdominal pain	Yes/No	Binary
Palpable liver	Yes/No	Binary
Haematocrit	Haematocrit value [%]	Continuous
Platelet count	Platelet count [cells per mm^3]	Continuous

5.2.3 Statistical analysis

As described in detail in Chapter 2, logistic regression was the statistical model of choice for all univariate and multivariable analyses. All candidate predictors were included in both univariate and multivariable analyses. The univariate analysis was based on the complete-case dataset, whereas I used multiple imputation for the multivariable analyses.

Details regarding the calculation of multivariable analyses and the chosen strategy for the development of the prognostic model for DSS are described in Chapter 2. Linearity assessment was performed for all continuous variables. Furthermore, I tested for possible interactions between gender and day of illness at enrolment, gender and all other covariates, and day of illness with all other covariates. Regarding model validation, I used a) 10-times 10-fold cross-validation and b) temporal validation, where the original dataset was split into a training set including 1663 patients enrolled before 2008, and a test set including 638 patients enrolled from 2008 onwards. For multiple imputation analysis, I followed current recommendations on how to perform these analyses on multiple imputation datasets (White et al., 2011) as detailed in Chapter 2 (Sections 2.2.3).

5.3 Results

5.3.1 General description

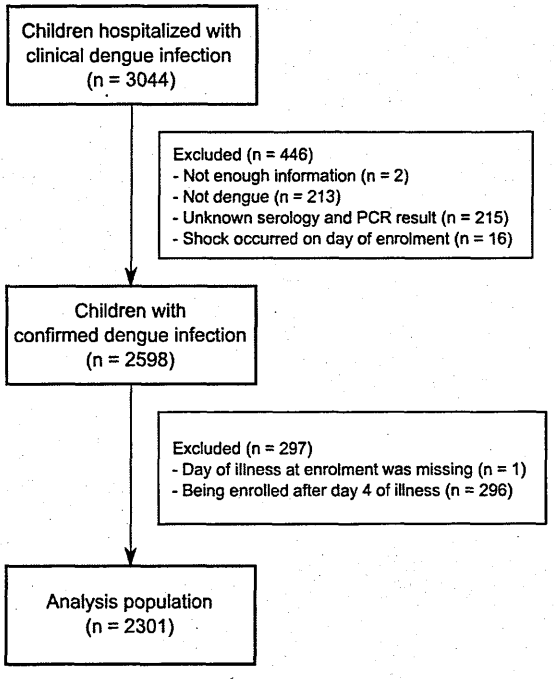

Figure 5.1. Flow-chart of the analysis for DSS development.

A total of 3044 children were enrolled into the MD cohort and 2598 of them had a laboratory confirmed dengue diagnosis. Amongst these, 2301 patients were enrolled on day 4 of illness or earlier and formed the primary analysis population (Figure 5.1).

Baseline characteristics

Characteristics of study participants at enrolment are summarized in Table 5.3. There were more males than females enrolled in this study and the median age was 12 years (IQR 10-13 years, range 5-16 years). Almost all patients were admitted to the Hospital for Tropical Diseases (HTD) within the first 4 days of illness (2411/2598, 94%). Most patients were then enrolled into the study within two days from hospital admission (67% on admission, 26% after 1 day, 6% after 2 days). The delay in enrolment of patient into the study may be explained by the fact that most study doctors were also treating doctors at the ward and extremely busy, especially during the dengue season. Another explanation is the fact that 206 (24% of 852 patients who were enrolled late) patients were admitted

on Saturday and Sunday, and therefore be enrolled one or two days later as study patients were only enrolled on weekdays.

As expected, most patients still had fever at enrolment (96% patients had body temperature $\geq 38°C$). In general, haemodynamic parameters including pulse rate and systolic BP were still within the normal range. There was one case with a systolic BP of 199 mmHg; however, this high blood pressure may not be attributable to dengue infection as the patient also had an underlying congenital heart disease.

At the time of enrolment, platelet values were quite low, whereas haematocrit values were slightly higher than the normal range. In patients whose follow-up platelet count after 14 days of illness was available, the platelet counts at enrolment were decreased by a median of 59% (IQR, 43% - 69%) compare to follow-up values; whereas, the level of haemoconcentration at enrolment was minimal (median 3%, IQR -3% to 10%).

Most cases were infected with DENV-1 and DENV-2. Amongst cases whose serology status could be determined, many (79%, 1601/2030) were secondary infections.

Table 5.3. Characteristics of participants at study enrolment.

Characteristics	Patients with dengue (n = 2598)			Patients with dengue before day 5 (n = 2301)		
	n	Summary statistics		n	Summary statistics	
Age [years]	2597	12	(10-13)	2300	12	(10-13)
Gender: Female	2598	1068	(41)	2301	939	(41)
Weight [kg]	2592	35.0	(27.0-42.0)	2296	34.5	(27.0-42.0)
Day of illness at enrolment	2597			2301		
- 1		12	(< 1)		12	(<1)
- 2		351	(13)		351	(15)
- 3		908	(35)		908	(39)
- 4		1030	(40)		1030	(45)
- 5		286	(11)		0	(0)
- 6		9	(<1)		0	(0)
- 7		1	(<1)		0	(0)
Days from admission to enrolment	2596			2299		
- 0		1746	(67)		1564	(68)
- 1		682	(26)		603	(26)
- 2		143	(6)		117	(5)
- 3		23	(1)		15	(1)
- 4		2	(<1)		0	(0)
History of headache: Yes	2597	1853	(71)	2301	1627	(71)
History of muscle pain: Yes	2586	584	(23)	2290	507	(22)
History of tiredness: Yes	2594	2201	(85)	2298	1936	(84)
History of diarrhea: Yes	2591	238	(9)	2295	196	(9)
History of cough: Yes	2594	398	(15)	2298	351	(15)
History of vomiting: Yes	2591	940	(36)	2295	832	(36)
Temperature [°C]	2596	39.0	(38.5-39.5)	2299	39.0	(38.5-39.5)
Temperature $\geq$ 38°C	2596	2492	(96)	2299	2210	(96)
Tourniquet test	2582			2289		
- Negative		1279	(49)		1164	(51)
- Equivocal		532	(21)		458	(20)
- Positive		771	(30)		667	(29)
Pulse [per min]	2593	100	(100-120)	2297	100	(100-116)
Systolic BP [mmHg]	2595	90	(90-100)	2298	93	(90-100)
Bleeding	2537			2244		
- None		1759	(69)		1615	(72)
- Skin only		578	(23)		471	(21)
- Mucosal		200	(8)		158	(7)
Abdominal pain: Yes	2586	556	(22)	2291	465	(20)

Palpable liver: Yes	2575	256	(10)	2279	217	(10)
Liver size below the costal margin[cm]	256	1	(1-2)	217	1	(1-2)
HCT [%]	2554	39.6	(37.3-42.1)	2259	39.5	(37.2-42.0)
Haemoconcentration [%][a]	2553	2.6	(-3.1-9.7)	2258	2.0	(-3.3-9.0)
PLT [1000 cells/mm^3]	2553	129	(92-174)	2258	134	(97-178)
PLT change [%][b]	1009	-58.6	(-69.6- -43.4)	894	-56.7	(-67.6- -41.9)
Serotype	2430			2152		
- DENV-1		1116	(46)		956	(44)
- DENV-2		583	(24)		553	(26)
- DENV-3		219	(9)		195	(9)
- DENV-4		176	(7)		169	(8)
- Mixed		9	(<1)		8	(<1)
- Negative		327	(13)		271	(13)
Immune status	2567			2271		
- Primary		141	(5)		114	(5)
- Possible primary		288	(11)		271	(12)
- Secondary		1601	(62)		1419	(63)
- Unclassifiable		537	(21)		467	(21)

Summary statistics are median (IQR) for continuous variables and frequency (%) for categorical variables.

[a] *Haemoconcentration was defined as the percentage change in the haematocrit value at enrolment compared to normal haematocrit (follow-up value or early value or population value).*

[b] *Comparing to follow-up value.*

Abbreviations: IQR = interquartile range, HCT = haematocrit, PLT = platelet count, DENV = dengue virus.

Clinical outcomes

During hospitalization, 200 cases (8%) were referred to the PICU for more intense monitoring and management (Table 5.4). Among 2598 patients with dengue, 156 (6%) patients developed DSS (55/156 or 35% of them were females) and there was no systematic time trend in the incidence of DSS over the study period (linear trend test: p value was 0.57). Even though DSS occurred on any day from day 3 to 8 of illness, the most critical period was from day 4 to day 6 (90% of DSS cases). In most cases, DSS happened within 3 days from enrolment (94%).

New bleeding occurred in 42% of cases, but in only 11% of these cases (120/1071) was the bleeding site mucosal. Among cases with mucosal bleeding, the most frequent bleeding sites were nose (66/120) and gum (35/120), with less frequent sites being gas-

trointestinal (16/120) and vaginal (5/120). Haemorrhage into the spinal cord occurred in 1 case.

The platelet nadir commonly occurred around day 6 with a median nadir value of 64,000 (IQR: 41,000-98,000) cells/mm^3. Many patients also had their highest haematocrit value on the same day with a median level of 44% (IQR 41%-47%) and a corresponding median maximum haemoconcentration of 13% (IQR 6% – 22%).

In general, the distributions of baseline characteristics and clinical outcomes in cases enrolled before day 5 were similar to those of all patients with dengue (Table 5.3 & 5.4).

Assessment of missing values

Amongst cases enrolled before day 5 of illness, 7% (171/2301) of participants had at least one missing value in one or more candidate predictors. The number of missing values per individual ranged from 0 to 6. HCT and PLT were the two most frequently missing predictors with 2% missing values.

The absence of several variables was found to be related to observed values of other variables: Tourniquet test results tended to be more frequently missing in cases who did not report tiredness, and the liver size below the costal margin was more often missing in cases enrolled earlier and cases with lower HCT values at enrolment.

As performing a complete case analyses would require ignoring 7% of data, and the MCAR assumption might be untenable in this situation, further multivariable analyses were based on imputed data sets using multiple imputation.

As described in detailed in Section 2.2.3 of Chapter 2, I created 20 imputed datasets using the MICE algorithm. Plots of the mean and variance of the imputations per stream by the iteration number suggested that the MICE algorithm had converged, as the variance between imputation streams was no larger than the variance within each stream without any observable trends. The imputed data were also reasonable as their values and distributions were similar to observed data.

Table 5.4. Clinical outcomes of study participants during hospitalization.

Characteristics	Patients with dengue (n = 2598)			Patients with dengue before day 5 (n = 2301)		
	n	Summary statistics		n	Summary statistics	
DSS: Yes	2598	156	(6)	2301	143	(6)
Day of illness at shock	156			143		
- 3		2	(1)		2	(1)
- 4		33	(21)		33	(23)
- 5		70	(45)		70	(49)
- 6		38	(24)		28	(20)
- 7		11	(7)		8	(6)
- 8		2	(1)		2	(1)
Days from enrolment to shock	156			143		
- 1		75	(48)		65	(45)
- 2		49	(31)		46	(32)
- 3		23	(15)		23	(16)
- 4		8	(5)		8	(6)
- 5		1	(1)		1	(1)
Referred to PICU: Yes	2598	200	(8)	2301	179	(8)
Bleeding during hospitalization	2583			2288		
- No		1508	(58)		1333	(58)
- Skin only		951	(37)		842	(37)
- Mucosal		120	(5)		110	(5)
- Other		4	(<1)		3	(<1)
Received IV fluid: Yes	2597	1859	(72)	2300	1664	(72)
Total volume of IV fluids [ml/kg]	1850			1657		
- Patients without DSS	1710	20.8	(14.3-33.3)	1528	21.7	(14.3-35.7)
- Patients with DSS	140	129.1	(98.8-164.5)	129	129.1	(100.0-165.0)
Platelet nadir [1000 cells/mm^3]	2569	64	(41-98)	2279	65	(41-99)
Day of platelet nadir	2569	6	(5-7)	2279	6	(5-7)
Maximum haematocrit [%]	2573	44	(41-47)	2283	44	(41-47)
Day of maximum haematocrit	2573	6	(5-6)	2283	5	(4-6)
Overall haemoconcentration [%]	2572	13	(6-22)	2282	13	(6-22)

Summary statistics are median (IQR) for continuous variables and frequency (%) for categorical variables. Total volume of IV fluids for patients with DSS included fluid given after development of DSS.
Abbreviations: DSS = dengue shock syndrome, IV = intravenous.

5.3.2 Risk factors of DSS

Univariate analysis

In univariate analyses, significant risk factors for developing DSS were male gender, a history of vomiting, higher temperature, abdominal pain, a palpable liver and lower platelet counts (Table 5.5). Regarding immune status, no case with a definite primary dengue infection developed DSS whereas 84/1419 (6%) cases with definite secondary infections progressed to DSS.

Table 5.5. Univariate effect of candidate predictors on the development of DSS amongst cases enrolled before day 5 of illness (n = 2301).

Covariate	Shock (n = 143)			No shock (n = 2158)			OR	(95% CI)	p value
	n	Summary statistics		n	Summary statistics				
Age [+1 year]	143	11	(10-13)	2157	12	(10-13)	0.97	(0.90, 1.05)	0.41
Gender: Female	143	47	(33)	2158	892	(41)	0.69	(0.48, 0.99)	0.04
Weight [+1 kg]	143	33	(27-40)	2153	35	(27-42)	0.99	(0.98, 1.01)	0.44
Day of illness at enrolment	143	3	(3-4)	2158	3	(3-4)	0.94	(0.75, 1.18)	0.58
History of tired: Yes	143	120	(84)	2155	1816	(84)	0.97	(0.63, 1.58)	0.91
History of vomiting: Yes	141	78	(55)	2154	754	(35)	2.30	(1.63, 3.25)	<0.01
Tourniquet test	140			2149					0.89
- Negative		69	(49)		1095	(51)	1.00		
- Equivocal		30	(21)		428	(20)	1.11	(0.71, 1.72)	
- Positive		41	(30)		626	(29)	1.04	(0.69, 1.54)	
Temperature [+1°C]	142	39.0	(38.6-39.5)	2157	39.0	(38.5-39.5)	1.33	(1.04, 1.71)	0.02
Pulse [+10 per min]	142	100	(100-120)	2155	100	(100-116)	1.07	(0.91, 1.27)	0.40
Systolic BP [+10 mmHg]	142	90	(90-100)	2156	95	(90-100)	1.01	(0.81, 1.23)	0.92
Mucosal bleeding: Yes	141	12	(9)	2142	146	(7)	1.27	(0.65, 2.26)	0.46
Abdominal pain: Yes	142	41	(29)	2149	424	(20)	1.65	(1.12, 2.39)	0.01
Palpable liver: Yes	143	28	(20)	2136	189	(9)	2.51	(1.59, 3.84)	<0.01
HCT [+1 %]	139	39.4	(38.0-42.6)	2120	39.5	(37.2-42.0)	1.03	(0.99, 1.07)	0.18
PLT [+10000 cells/mm^3]	139	104000	(76400-150000)	2119	136000	(99000-180000)	0.92	(0.89, 0.95)	<0.01
Serotype	133			1748					0.08
- DENV-1		65	(49)		891	(51)	1.00		
- DENV-2		46	(34)		507	(29)	1.24	(0.84, 1.84)	
- DENV-3		7	(5)		188	(11)	0.51	(0.21, 1.06)	
- DENV-4		13	(10)		156	(9)	1.14	(0.59, 2.06)	
- Mixed		2	(2)		6	(<1)	4.57	(0.66, 20.29)	
Immune status	129			2142					<0.01
- Secondary dengue		84	(65)		1335	(62)	1.00		
- Primary dengue		0	(0)		114	(5)	0.07	(0.00, 0.48)	
- Possible primary		45	(35)		226	(11)	3.17	(2.14, 4.65)	
- Unclassifiable		0	(0)		467	(22)	0.01	(0.00, 0.12)	

Summary statistics are median (interquartile range) for continuous variables and frequency (percentage) for categorical variables. Analyses were based on observations where the respective covariate was non-missing (complete cases). OR, CI and p-value for immune status were calculated based on penalized maximum likelihood (Firth's correction).

Abbreviations: HCT = haematocrit, PLT = platelet count, DENV = dengue virus, BP = blood pressure, OR = odds ratio, CI = confidence interval.

Multivariable analysis

The linearity assessment suggested that linear terms were sufficient for all continuous candidate predictors except that there was some indication of non-linearity for the effect of age on the development of DSS, as displayed in Figure 5.2. However, the displayed non-linearity for age was not strong and linearity tests mostly did not reach statistical significance. There was also no evidence of any interactions between gender or day of illness with any other variables (Table 5.6). Thus, no non-linear terms or interactions were added to the pre-defined multivariable model.

Table 5.6. *Linearity and additivity tests in the pre-defined multivariable logistic regression model for the development of DSS using complete-case and multiple imputation analyses.*

	Complete case analysis			Multiple imputation
	Deviance	**df**	**p value**	**p value**
Linearity tests (compared to a quadratic function)				
Age	3.94	1	0.05	0.06
Weight	0.88	1	0.35	0.37
Temperature	0.10	1	0.75	0.75
Pulse	0.35	1	0.56	0.52
Systolic BP	0.01	1	0.91	0.76
HCT	0.50	1	0.48	0.50
PLT	1.13	1	0.29	0.25
Linearity tests (compared to a natural cubic spline with 4 degrees of freedom)				
Age	5.30	3	0.15	0.12
Weight	6.98	3	0.07	0.16
Temperature	5.12	3	0.16	0.14
Pulse	1.47	3	0.69	0.59
Systolic BP	3.17	3	0.37	0.55
HCT	2.18	3	0.54	0.64
PLT	2.90	3	0.41	0.39
Additivity assessment (interaction tests)				
Gender vs. others	12.39	15	0.65	0.80
Day of illness vs. others	13.82	15	0.54	0.30
Gender vs. age	0.01	1	0.91	0.87

Abbreviations: BP = blood pressure, HCT = haematocrit, PLT = platelet count.

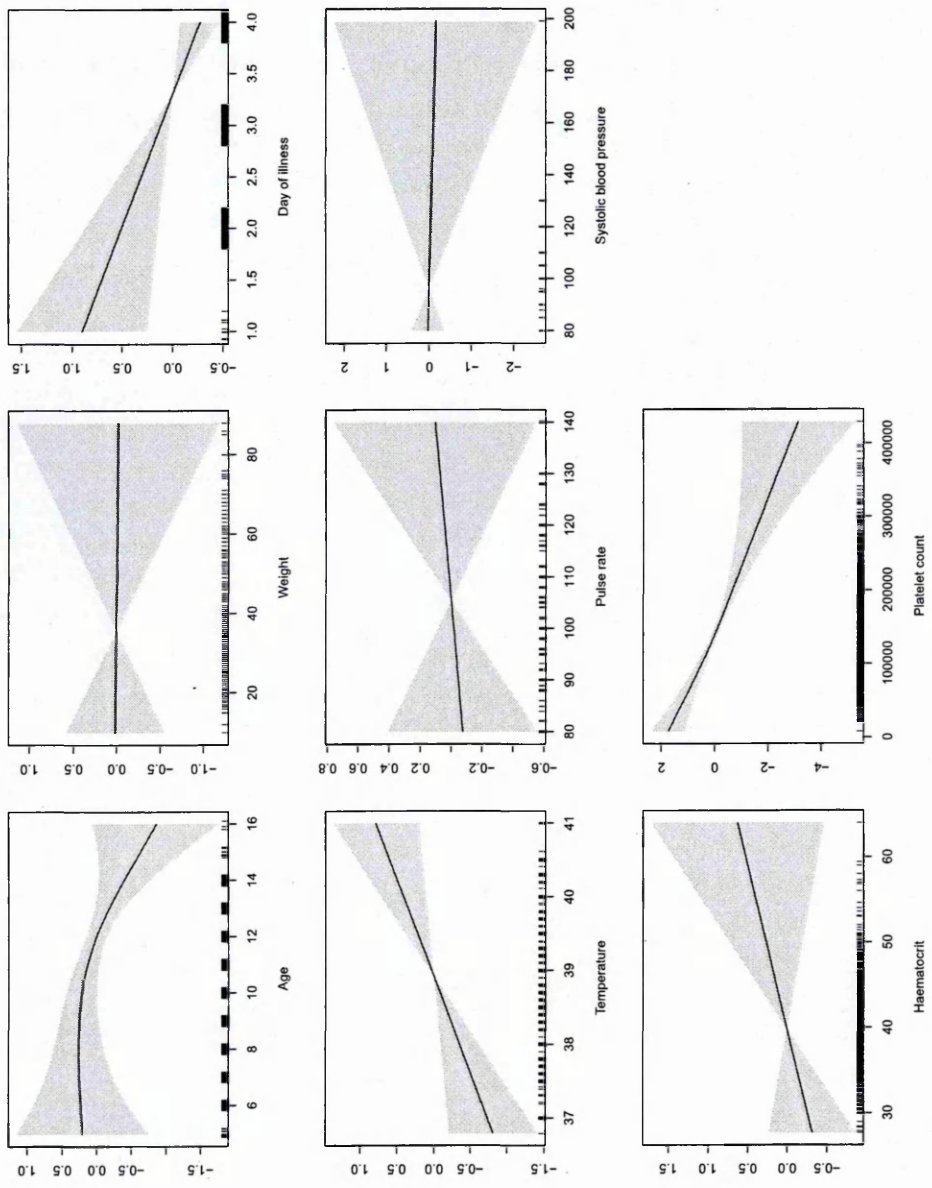

Figure 5.2. *Plots of estimated component smooth functions of a generalized addictive model (GAM) fit for development of DSS with continuous covariates modelled using natural cubic spline functions and integrated smoothness estimation. Solid lines correspond to spline functions estimated by GAM; dashed lines correspond to point-wise 95% confidence intervals of the estimated values. This plot was based on a complete case analysis.*

In the multivariable analysis using all pre-defined candidate predictors, identified risk factors for developing DSS included male gender, enrolment at an earlier day of illness, vomiting, higher temperature, a palpable liver, and a lower platelet count. This result was consistent between the complete-case analysis and the analysis based on multiple imputation (Table 5.7). In both univariate and multivariable analyses, there was evidence of a protective effect of being female (Table 5.5 & 5.7). However, this effect no longer reached statistical significance when I re-did the univariate and multivariable analyses on all 2598 patients with dengue infection regardless of the day of enrolment (Table 5.10).

Table 5.7. *Adjusted effect of candidate predictors on the development of DSS amongst cases enrolled before day 5 of illness in complete-case and multiple imputation analyses (n = 2301).*

Covariate	Complete-case analysis			Multiple imputation analysis		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.92	(0.83, 1.01)	0.09	0.94	(0.85, 1.04)	0.22
Gender: Females	0.64	(0.43, 0.94)	0.02	0.65	(0.44, 0.94)	0.02
Weight [+1 kg]	1.00	(0.98, 1.02)	0.91	0.99	(0.97, 1.02)	0.61
Day of illness	0.67	(0.51, 0.88)	<0.01	0.68	(0.52, 0.88)	<0.01
History of tiredness: Yes	0.92	(0.56, 1.59)	0.76	0.88	(0.54, 1.43)	0.61
History of vomiting: Yes	2.17	(1.51, 3.14)	<0.01	2.19	(1.53, 3.13)	<0.01
Tourniquet test			0.46			0.47
- Negative	1.00			1.00		
- Equivocal	1.05	(0.64, 1.66)		1.11	(0.70, 1.76)	
- Positive	0.78	(0.50, 1.21)		0.82	(0.54, 1.25)	
Temperature [+1°C]	1.45	(1.10, 1.91)	<0.01	1.39	(1.07, 1.82)	0.02
Pulse [+10 per min]	1.03	(0.85, 1.23)	0.79	1.03	(0.86, 1.23)	0.79
Systolic BP [+10 mmHg]	0.99	(0.77, 1.23)	0.90	1.01	(0.81, 1.25)	0.94
Mucosal bleeding: Yes	1.09	(0.52, 2.06)	0.81	1.17	(0.61, 2.25)	0.63
Abdominal pain: Yes	1.12	(0.69, 1.79)	0.63	1.05	(0.66, 1.67)	0.83
Palpable liver: Yes	1.74	(0.99, 2.98)	0.05	1.74	(1.02, 2.98)	0.04
HCT [+1 %]	1.02	(0.98, 1.07)	0.28	1.03	(0.98, 1.07)	0.23
PLT [+10,000 cells/mm^3]	0.89	(0.85, 0.92)	<0.01	0.89	(0.86, 0.93)	<0.01

Abbreviations: BP = blood pressure, HCT = haematocrit, PLT = platelet count, OR = odds ratio, CI = confidence interval.

5.3.3 Prediction models

Age, gender, day of illness, history of vomiting, temperature, palpable liver and platelet count were retained in the logistic regression model with stepwise variable selection (Table 5.8). The same predictors and similar effect sizes were chosen by the complete-case and

the multiple imputation analysis.

Table 5.8. *Reduced model for the development of DSS with variable selection (complete-case and multiple imputation).*

Covariate	Complete-case analysis			Multiple imputation analysis		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.92	(0.85, 1.00)	0.04	0.93	(0.86, 1.01)	0.09
Gender: Female	0.63	(0.43, 0.92)	0.02	0.64	(0.44, 0.92)	0.02
Day of illness	0.68	(0.52, 0.90)	<0.01	0.69	(0.53, 0.89)	0.01
History of vomiting: Yes	2.19	(1.53, 3.16)	<0.01	2.19	(1.54, 3.11)	<0.01
Temperature [+1°C]	1.43	(1.10, 1.86)	<0.01	1.36	(1.05, 1.75)	0.02
Palpable liver: Yes	1.78	(1.08, 2.83)	0.02	1.76	(1.11, 2.80)	0.02
PLT [+10,000 cells/mm^3]	0.89	(0.85, 0.92)	<0.01	0.89	(0.86, 0.93)	<0.01

95% confidence intervals and p values do not take into account the uncertainty of model selection.
Abbreviations: PLT = platelet count, OR = odds ratio, CI = confidence interval.

In both temporal and internal validation, CART was inferior whereas all the other models had similar performance. The reduced logistic regression models with variable selection performed comparably to the full logistic model, and there were no big differences in performance between models with and without gender as a covariate. In internal validation, all models (except for CART) had moderate AUC and good calibration. However, there was an indication of over-fitting in temporal validation when discrimination of all models was substantially lower and miscalibration was apparent for logistic regression models. As the incidence of DSS between training and test sets were similar (91/1552 or 6% and 43/634 or 7% for training and test set, respectively), the poor model performance in temporal validation may be explained by the relatively low effective sample size in both the training and the test set which could have led to over-fitting and unstable coefficient estimates in the training set and imprecise estimates of performance in the test set. Another explanation could be the existence of interactions between some risk factors (for example age, gender, or day of illness at enrolment) and time, which could be investigated by appropriate interaction tests in the statistical model using data from all patients. The results were similar between complete case and multiple imputation analyses (Table 5.9).

Applying the reduced logistic regression model (complete case analysis, including gender as a covariate) on the original dataset yielded a skewed distribution of subject-specific predicted risks (median 0.045, IQR 0.026 – 0.079). Figure 5.3 displays the number of

true positive and false positive cases depending on the chosen risk threshold for classifying subjects as likely to progress to DSS or not. For a low risk threshold, the number of false positive cases is quite high. For example, at a risk threshold of 5%, 108/134 (81%) of cases with DSS would be correctly classified; however, the number of false positive cases would be eight times higher (894 cases). For a higher risk threshold, the number of false positive cases is decreased at the cost of missing true positive cases. For example, at a risk threshold of 20%, there are only 46 false positive cases but only 17/134 (13%) of cases with DSS would be detected by the model.

As the relatively low incidence of DSS and the moderate performance of the prediction model jointly indicate that the presented prediction model is of limited clinical usefulness, I decided not to simplify the model for clinical use, e.g. to create a score chart or nomogram.

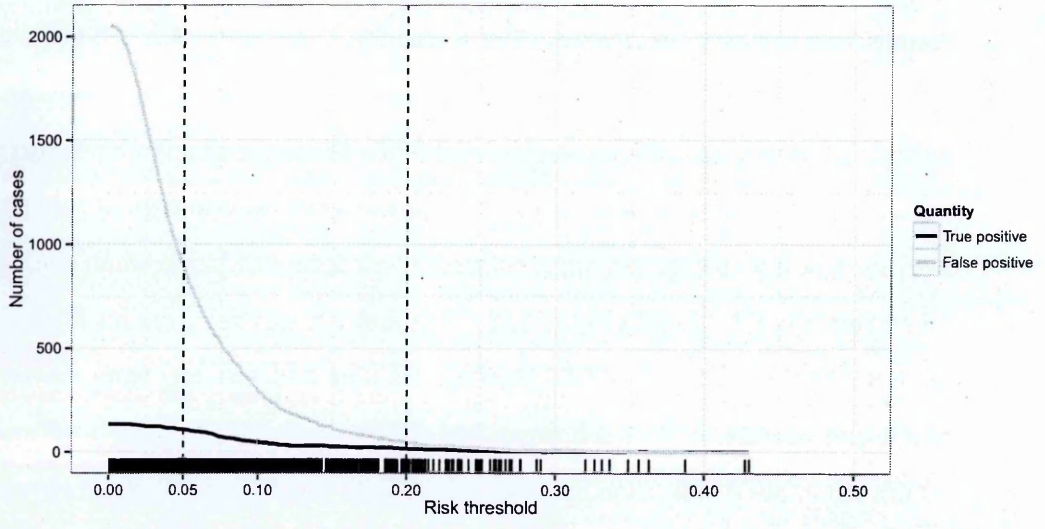

Figure 5.3. *The number of true positive and false positive cases when the reduced logistic regression model (including gender and based on complete-case analysis) is applied on the original dataset using different risk thresholds for classification. Rugs at the bottom correspond to the distribution of predicted risks. The two vertical lines correspond to risk thresholds of 5% and 20%.*

Table 5.9. Performance of different prediction models for development of DSS based on complete-case and multiple imputation analysis.

Measure	Logistic 1	Logistic 2	Logistic 3	Logistic 4	Lasso	GAM	CART	Boosting
Complete-case analysis								
Temporal validation								
Brier score	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
AUC	0.64	0.65	0.63	0.63	0.65	0.64	0.58	0.65
(95% CI)	(0.56, 0.72)	(0.56, 0.73)	(0.54, 0.71)	(0.54, 0.71)	(0.57, 0.73)	(0.56, 0.72)	(0.50, 0.66)	(0.56, 0.73)
Calibration in-the-large	0.33	0.35	0.33	0.35	0.29	0.34	0.14	0.13
(95% CI)	(-0.00, 0.63)	(0.01, 0.65)	(-0.00, 0.64)	(0.02, 0.66)	(-0.05, 0.58)	(0.01, 0.65)	(-0.20, 0.45)	(-0.19, 0.44)
Calibration slope	0.52	0.57	0.48	0.52	0.79	0.53	0.29	0.75
(95% CI)	(0.18, 0.88)	(0.22, 0.93)	(0.14, 0.84)	(0.17, 0.88)	(0.31, 1.29)	(0.18, 0.88)	(-0.14, 0.67)	(0.19, 1.26)
Internal validation								
Brier score	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
AUC	0.70	0.71	0.70	0.70	0.70	0.69	0.57	0.69
Calibration in-the-large	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.12
Calibration slope	0.84	0.94	0.85	0.92	1.09	0.77	-0.07	1.15
Multiple imputation								
Temporal validation								
Brier score	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06
AUC	0.64	0.65	0.62	0.63	0.65	0.64	0.56	0.65
(95% CI)	(0.55, 0.72)	(0.56, 0.73)	(0.54, 0.71)	(0.54, 0.71)	(0.57, 0.73)	(0.55, 0.72)	(0.48, 0.65)	(0.57, 0.74)
Calibration in-the-large	0.33	0.31	0.33	0.31	0.26	0.32	0.12	0.10
(95% CI)	(0.01, 0.65)	(-0.01, 0.63)	(0.01, 0.65)	(-0.01, 0.63)	(-0.05, 0.58)	(0.00, 0.64)	(-0.22, 0.46)	(-0.22, 0.41)
Calibration slope	0.53	0.58	0.49	0.53	0.73	0.53	0.31	0.82
(95% CI)	(0.17, 0.88)	(0.21, 0.95)	(0.12, 0.85)	(0.16, 0.90)	(0.27, 1.19)	(0.17, 0.88)	(-0.19, 0.80)	(0.27, 1.37)
Internal validation								
Brier score	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
AUC	0.69	0.70	0.69	0.70	0.69	0.68	0.57	0.68
Calibration in-the-large	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.03	-0.10
Calibration slope	0.83	0.91	0.84	0.91	1.08	0.77	-0.03	1.12

Logistic 1: full logistic regression model including all pre-defined candidate risk factors; logistic 2: reduced model based on the full model including all pre-defined candidate factors and stepwise variable selection; logistic 3: full logistic regression model including all pre-defined candidate risk factors except for gender; logistic 4: reduced model based on the full model including all pre-defined candidate factors except for gender and stepwise variable selection.

5.4 Discussion

This chapter identified male gender, enrolment at an earlier day of illness, a history of vomiting, higher temperature, a palpable liver, and lower platelet counts at enrolment as risk factors for DSS amongst children hospitalized with dengue infection. Based on these identified factors, I developed prediction models for DSS with moderate performance but rather limited clinical usefulness.

The incidence of DSS in this study (6%) was lower compared to previous studies (Anders et al., 2011; Alexander et al., 2011; Giraldo et al., 2011; Gupta et al., 2011; Chuansumrit et al., 2010; Mena Lora et al., 2014) where this number varied from 10% to 20%. This could be because this cohort aimed to include patient who were hospitalized early (within 3-4 days of illness) whereas other studies assessed a more general population of patients hospitalized with dengue. Another explanation is the fact that patients who were very sick at hospital admission would not have been included in this study, as they are referred directly to the PICU rather than the general hospital ward where this study took place.

In agreement with the current literature, thrombocytopenia and clinical warning signs including vomiting and a palpable liver were identified as predictors of DSS in this analysis (World Health Organization, 2009; Huy et al., 2013b). In addition, higher temperature at enrolment was also independently associated with an elevated risk of DSS, which might be explained by the positive correlation between temperature and viral load (Tsai et al., 2013; Vaughn et al., 1997).

In this study, there was 114/2271 (5%) cases classified as primary dengue and none of them developed DSS later; whereas 45/271 (17%) of possible primary cases and 84/1419 (6%) of secondary cases developed DSS. However, this result should be interpreted with caution because the immune status classification was based only on IgG result and was expected to be imprecise in detecting secondary infection (Section 2.1.3 of Chapter 2). In an attempt to verify the classification used in this study, I also assessed immune status of participants using a new classification algorithm has been developing in our unit, which based on IgM/IgG ratio and allows cut-off value to vary over time. Amongst 628 cases whose immune status can be determined using both method, most cases with primary

(32/42, 76%) and possible primary infection (67/96, 70%) in the current system were also classified as primary in the new system; however, only 55% (162/297) secondary cases in the current system were classified as secondary in the new system.

Enrolment at an earlier day of illness was associated with a significantly higher risk of DSS in multivariable but not univariate analyses. However, this could be an artifact attributable to the adjustment for other clinical signs and symptoms at enrolment, especially platelet count, in the multivariable analysis. By adjusting for platelet count, the reported odds ratio corresponds to the comparison of two subjects who were enrolled on two consecutive days of illness but had the same platelet count on their respective days of enrolment. As platelet count is known to decrease over time during dengue illness (Dinh The et al., 2012), the subject enrolled earlier would have a lower platelet count relative to their day of enrolment and, as platelet count is strongly inversely associated with the risk of DSS development, this might explain the reported effect.

In this study, there was a significant relationship between male gender and a higher risk of developing DSS, which is in contradiction to evidence from previous epidemiologic studies (Huy et al., 2013b). However, this association was only significant in the primary study population of patients enrolled before day 5 of illness but not in all enrolled patients with confirmed dengue. Of note, in the DF cohort, females were more likely to be admitted on the day of DSS than males (% of cases admitted on the day of DSS were 49% for females and 41% for males, Chi-squared test p value was <0.01). As "severe" cases, who might develop DSS shortly after hospital admission, were underrepresented in the MD study, "severe" females might also be underrepresented which could explain our results. Further research is required to shed light on the role of gender in dengue infection.

Amongst all developed prediction models in this analysis, the reduced logistic regression model based on the full logistic regression model with all pre-defined candidate predictors and stepwise variable selection was the model with the best trade-off between transparency/simplicity and accuracy. Unfortunately, the clinical usefulness of this model nevertheless appears to be rather limited even though it achieved a moderate performance in both temporal and internal validation. To be useful in clinical practice, a prediction model would need to be able to correctly identify most subjects who subsequently develop DSS. However, as illustrated in Figure 5.3, this would mandate a very low risk threshold

which would imply that the number of true positives would be swamped by the much larger number of false positives.

As the current model was carefully developed based on a relatively large sample size, the limited usefulness of the derived prognostic models could indicate that readily available baseline characteristics and warning signs are in general insufficient for reliable prediction of DSS in hospitalized patients. Novel markers with higher predictive value might be required in order to achieve a better prediction model. However, to identify novel predictors, which are not routinely collected currently, and implement them in clinical practice might require a lot of time and effort. An alternative approach would be to try to incorporate longitudinal information of risk factors, which are often already available in clinical practice. This is examined in the next chapter.

5.5 Appendix

Table 5.10. *Unadjusted and adjusted effect of candidate predictors on the development of DSS amongst all patients with dengue (complete-case analysis amongst n=2598 subjects amongst which n=2186 had complete data).*

Covariate	Complete-case analysis			Multiple imputation analysis		
	OR	95% CI	p value	OR	95% CI	p value
Age [+1 year]	0.97	(0.90, 1.04)	0.34	0.91	(0.83, 1.01)	0.07
Gender: Female	0.77	(0.55, 1.07)	0.12	0.73	(0.51, 1.05)	0.09
Weight [+1 kg]	0.99	(0.98, 1.01)	0.34	1.00	(0.98, 1.02)	0.98
Day of illness	0.89	(0.74, 1.07)	0.21	0.65	(0.52, 0.81)	<0.01
History of tiredness: Yes	0.93	(0.61, 1.48)	0.76	0.87	(0.54, 1.46)	0.58
History of vomiting: Yes	2.34	(1.69, 3.26)	<0.01	2.17	(1.53, 3.09)	<0.01
Tourniquet test			0.89			0.58
- Negative	1.00			1.00		
- Equivocal	1.09	(0.71, 1.65)		1.03	(0.65, 1.60)	
- Positive	1.07	(0.73, 1.56)		0.82	(0.54, 1.24)	
Temperature [+1°C]	1.35	(1.06, 1.71)	0.01	1.45	(1.11, 1.88)	<0.01
Pulse [+10 per min]	1.08	(0.92, 1.26)	0.35	1.02	(0.85, 1.21)	0.86
Systolic BP [+10 mmHg]	1.03	(0.83, 1.24)	0.79	1.01	(0.80, 1.24)	0.94
Mucosal bleeding: Yes	1.10	(0.58, 1.91)	0.75	0.93	(0.47, 1.70)	0.82
Abdominal pain: Yes	1.70	(1.18, 2.40)	<0.01	1.20	(0.76, 1.85)	0.43
Palpable liver: Yes	2.42	(1.57, 3.62)	<0.01	1.70	(1.00, 2.83)	0.05
HCT [+1 %]	1.02	(0.98, 1.06)	0.24	1.02	(0.98, 1.07)	0.32
PLT [+10000 cells/mm3]	0.92	(0.89, 0.95)	<0.01	0.89	(0.85, 0.92)	<0.01

Chapter 6

Dynamic prognostic models in acute diseases

Summary

This chapter provides an overview of current approaches to developing and assessing dynamic prediction models. Differences between acute and chronic disease settings and their implications for statistical modelling are discussed. The chapter concludes with a case study which describes and compares several dynamic prediction models for the development of DSS in hospitalized dengue patients.

6.1 Introduction to dynamic prognostic models

As discussed in Chapter 1 (Section 1.3.1), dynamic prediction models allow predicting the future course of the disease at follow-up time points based on the updated longitudinal information. In many settings, such models provide more accurate predictions compared to baseline models (Lemeshow et al., 1988; Christensen et al., 1993; Hughes et al., 1992; Rué et al., 2001; Karp et al., 2004). Dynamic prediction may also be appealing for clinicians as it mimics the iteration of obtaining information and updating prognosis based on this new information, a task that physicians routinely do every day in clinical practice.

6.2 Modelling approaches to dynamic prediction models

A naive strategy to obtain a dynamic prediction is to apply a baseline model sequentially over time by simply plugging in the time-updated covariate values. Even though this strategy might work better than a traditional baseline model in some settings of chronic diseases (Karp et al., 2004), it is conceptually inappropriate as baseline models should only be used to provide predictions for future patients from the same time origin as that used in the model development (Hughes et al., 1992).

Let $Y(u)$ denote the event status of the outcome of interest at time u, i.e. $Y(u) = 1$ if the outcome occurred at or before time u and $Y(u) = 0$ otherwise, and let $Z(t)$ denote the value of the time-varying predictor variables at time t. Then the goal of dynamic prediction modelling, which is to predict the conditional probability of the event occurrence of Y up to a future time point u depending on the patient history Z up to the current time point t, can be written as $\pi(u|t) = P(Y(u) = 1|Z(s)$ for $s \leq t$ and $Y(t) = 0)$ (Van Houwelingen and Putter, 2012). I will schematically denote this conditional probability by $[Y|Z]$. Fundamentally, there are two ways to obtain this conditional probability: either one can model the conditional probability directly, or one can model the joint probability $[Y, Z]$ first and then get the quantity of interest from the joint model, i.e. $[Y|Z] = \frac{[Y,Z]}{[Z]}$. In the conditional approach, one can base the conditional probability of interest $\pi(u|t)$ on either (1) the complete history of the time-dependent covariates up to the current time t, or (2) a subset or some aspects of the history of time-dependent covariates up to time t. Even though approach (1) is often desired, approach (2) is easier to conduct and, if

the relevant aspects of the history of time-dependent covariates up to time t are chosen carefully, usually adequate (Pepe and Couper, 1997; Diggle et al., 2002).

6.2.1 Conditioning on the complete underlying history of the longitudinal process

One model that falls in this category is the Cox proportional hazards regression model with time-dependent covariates (Cox, 1972) that models the hazard rate of an event at time t, $\lambda(t)$, as follows:

$$\lambda(t|Z_i(t)) = \lambda_0(t)\exp\left(\beta_1 W_i + \beta_2 Z_i(t)\right)$$

where $\lambda_0(t)$ is the baseline hazard at time t, W_i are the time-fixed covariates of subject i, $Z_i(t)$ denotes the longitudinal time-dependent covariate values of subject i at time t (which can also include all observed past values, lagged values, or changes of the longitudinal process); and β_1, β_2 are the corresponding vectors of regression coefficients.

For parameter estimation, values of Z_i at all observed event times are required (Collett, 2003). As these values might not be available in practice where longitudinal variables are only collected at discrete time points, they have to be imputed, for example by using the "last observation carried forward" method or linear interpolation between consecutive observed values (Collett, 2003). Moreover, as only longitudinal values at observed event times are used in the model estimation stage, this approach discards a lot of information (Altman and De Stavola, 1994), especially when the frequency of events is low.

Based on this model, the dynamic prediction for the event status at the future time u given the current history at time t of subject i can be approximated by the probability of having an event at time u given that the subject is event-free at time t, which is defined as

$$Pr(Y_i(u) = 1|Y_i(t) = 0, W_i, Z_i(0), \ldots, Z_i(t)) = 1 - \frac{S_i(u)}{S_i(t)}$$

where $S_i(t)$ is the survival function of subject i at time t.

As survival function in time-dependent Cox proportional hazards regression model depends on all values of time-dependent covariates from baseline to the time point of interest, the right hand side of the above equation depends on future values of Z_i from

time t to time u, which are not available at time t. Therefore, dynamic prediction at a specific time point depends not only on the entire longitudinal covariate profile up to the current time t, but also on the future unobserved evolution of the longitudinal process up to time u. Dynamic predictions must thus rely on assumptions regarding the future development of the longitudinal markers, for example that they remain constant (Altman and De Stavola, 1994). This also leads to a conceptual difficulty when applying this type of model with internal time-dependent covariates, especially when the event is death, as the existence of a covariate value is contingent upon the survival of the patient up to that time point (Fisher and Lin, 1999). While these are major issues, time-dependent Cox regression models are also easy to fit in standard statistical software, for example using the coxph function in the R library survival, and have been applied to develop dynamic prediction models in various settings (Christensen et al., 1993; Karp et al., 2004; Hartmann et al., 2012).

6.2.2 Conditioning on some aspects of the history of the longitudinal process

In this approach, certain aspects of the history of the longitudinal process which are considered most relevant to outcome prediction, for example all observed past values, the current value, previous values or the change in these values, are used to obtained dynamic predictions. Each of these aspects could be modelled using either the person-interval or the partly conditional modelling approach outlined below.

Person-interval approach

In this approach, the follow-up time of each participant is split into intervals and then information regarding covariate values at or before the interval, and outcome occurrence during each interval are used for parameter estimation. Different models have been proposed depending on how person-intervals are defined. One splitting strategy is to divide individual follow-up times into short, distinct intervals of equal length (Wu and Ware, 1979; Cupples et al., 1988; Ruttimann and Pollack, 1991; Hughes et al., 1992). Starting points of the interval can be defined either as times when repeated measurements are recorded (Wu and Ware, 1979; Cupples et al., 1988; Ruttimann and Pollack, 1991;

Hughes et al., 1992) or as consecutive starting points at equal distance according to the pre-defined interval length (Hughes et al., 1992). Another splitting strategy is to divide individual follow-up time into intervals of equal length that start at a pre-defined time point, landmark time t_{LM}, and stop at $t_{LM} + w$ where w is the pre-defined window of prediction (Van Houwelingen and Putter, 2012). In this landmarking approach, the derived intervals from the same individual may be distinct or may overlap, depending on whether the prediction window is smaller or larger than the gap between landmark points. Another splitting strategy which might result in intervals of variable length is to define the intervals to start at measurement time points and to end at the next consecutive measurement time points (Murtaugh et al., 1994). A simple illustration of these splitting strategies is presented in Figure 6.1. When biomarker values are measured at regular time points and the length of the interval is set equal to the gap between consecutive measurement points, all of these splitting strategies lead to the same intervals.

Given the interval, logistic regression or Cox regression models can be used to model the outcome of interest within that interval conditional on past covariate values. In principle, the relationship between covariates and outcome can vary across different intervals, and one can also use all observed longitudinal values up to the current interval resulting in a very general model as proposed by Wu and Ware (1979):

$$logit\{Pr(Y_{i,t} = 1|W_i, \mathbb{Z}_{i,t})\} = \beta_0^{(t)} + \beta_1^{(t)} W_i + \sum_{j=0}^{t} \beta_2^{(t,j)} Z_{i,j}$$

where $Y_{i,t}$ is the outcome of subject i during interval t, W_i are fixed time-independent covariates, $Z_{i,t}$ is the observed value of repeated measurement of subject i at the beginning of interval t , $\mathbb{Z}_{i,t}$ is the collection of all repeated measurements $Z_{i,j}$ of subject i until time t ($j \leq t$), and $\beta_0, \beta_1, \beta_2$ are the corresponding regression coefficients.

In this model, two assumptions are made: (1) both current measurements and earlier measurements of potential risk factors contribute to the overall linear predictor of the risk score in an additive way and (2) the same intervals are selected for each participant, which essentially requires that all patients be assessed at the same time points. This approach uses all observed repeated measurements but does not require rich longitudinal data. However, two main drawbacks of this approach are that it (1) involves a large number of

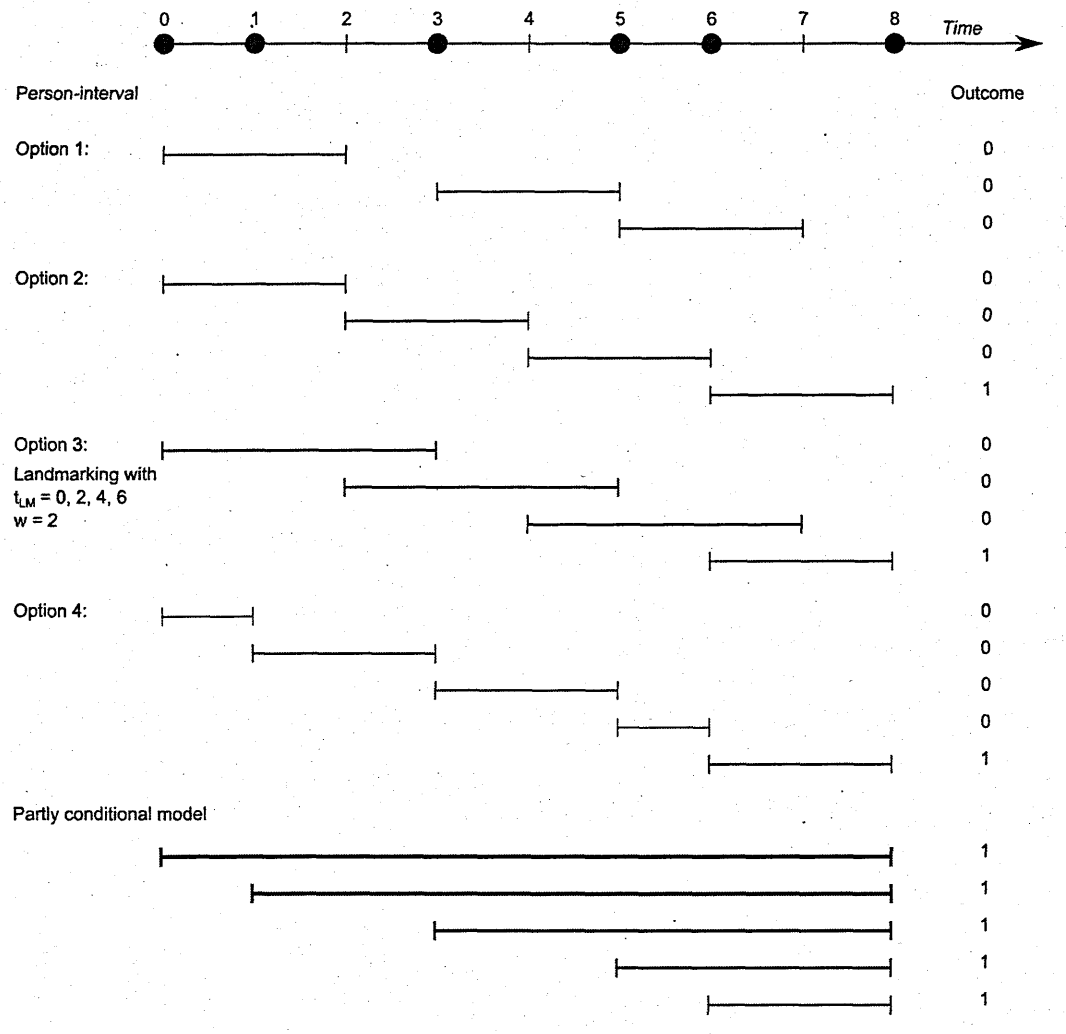

Figure 6.1. *Illustration of different strategies to define intervals in person-interval and partly conditional modelling approaches. Red dots correspond to time points when repeated measurements are recorded; the black dot corresponds to the time when the event occurs.*

parameters, which might be difficult to estimate with a limited sample size, and that it (2) provides different logistic models with different numbers of parameters for each time point of assessment.

Additional assumptions are required to reduce the number of parameters in the approach described above, for example one can assume regression coefficients to be the same in all intervals (Ruttimann and Pollack, 1991) or time-varying according to simple linear functions only (Wu and Ware, 1979). Ruttimann and Pollack (1991) further reduced the number of parameters by applying likelihood ratio tests in a stepwise backwards procedure. By assuming that the relationship between covariates and outcome is independent of the time interval, data from all intervals can be pooled together as if the

information recorded in each interval is new observation (Cupples et al., 1988; Hughes et al., 1992; Murtaugh et al., 1994). Further assuming that only the current values of the longitudinal markers are relevant to outcome prediction results in a very simple and straightforward model which is just a standard logistic regression or Cox regression model applied to this pooled dataset

$$logit\{Pr(Y_{i,t} = 1|W_i, Z_{i,t})\} = \beta_0 + \beta_1 W_i + \beta_2 Z_{i,t}$$

$$\lambda(u|W_i, Z_{i,t}) = \lambda_0(u) \exp\{\beta_1 W_i + \beta_2 Z_{i,t}\}$$

A clear strength of this approach is that it is both easy to interpret and "dynamic": it is easy because it uses the same model at each time point and the dynamic updating can easily be implemented by plugging the current values of the risk factors into the model. However, its assumptions may be implausible in the setting of rapidly progressive diseases or if the prediction intervals are long.

A more recent approach in this category is the landmarking method, which proposes to fit standard models to individuals still at risk at the landmark time point t_{LM} and applies administrative censoring at the time horizon $t_{LM} + w$ (Van Houwelingen and Putter, 2012). Specifically, the main idea is to fit a standard Cox model to a big dataset that stacks all at-risk datasets from each landmark point. Based on this, we can get an approximate risk prediction for an individual at a certain time horizon $u = t_{LM} + w$ given their risk factors at landmark point t_{LM}. This landmarking approach is somewhat similar to the above approaches except that it allows baseline hazards and/or regression coefficients to depend on the landmark point by fitting separate models for different landmark points. The approach uses either a Cox model stratified by the landmark points (if different baseline hazards for each landmark point are desired) or an analysis involving delayed entry (if a common baseline hazard is desired). Furthermore, as the person-intervals in the landmarking approach can be overlapping, naive standard error of estimated regression coefficients will be too narrow, and therefore, they have to be corrected, for example by using sandwich estimators of the covariance matrix (Van Houwelingen and Putter, 2012).

Obtaining dynamic predictions from all of the above models is relatively straightforward. Specifically, they can be obtained by either plugging-in estimated parameters and individual covariate information into the logistic model, or by using the formula

$$\begin{aligned}\pi_i(u|t) &= 1 - \frac{S_i(u)}{S_i(t)} \\ &= 1 - \exp\left\{-\int_t^u \lambda_0(s)\exp(\beta_1 W_i + \beta_2 Z_{i,t})ds\right\} \\ &= 1 - \exp\left\{-\exp(\beta_1 W_i + \beta_2 Z_{i,t})(H_0(u) - H_0(t))\right\}\end{aligned}$$

where $H_0(t)$ is the cumulative hazard function at time t for a Cox model.

As these models, except for landmarking, involve splitting the follow-up time into relative short periods, these models might only be appropriate for short-term prediction within the pre-defined interval and extrapolation beyond that interval can lead to misleading results (Hughes et al., 1992). In the landmarking approach, short-term or long-term prediction can be obtained by adjusting the prediction window. Several dynamic models using this person-interval approach have been derived and claimed to perform better than baseline prediction models (Ruttimann and Pollack, 1991; Hughes et al., 1992; Murtaugh et al., 1994). As an example, the updated natural history prediction model for primary biliary cirrhosis was developed based on this approach (Murtaugh et al., 1994) and is available as an online tool for use in clinical practice (Mayo Clinic, 2015).

Partly conditional models approach

Unlike the person-interval approach, partly conditional models always involve splitting follow-up time into overlapping intervals which start at each repeated measurement point and last till the end of the follow-up time (Figure 6.1). In addition, partly conditional survival models as proposed by Zheng and Heagerty (Zheng and Heagerty, 2005) also require to reset the time clock to zero at the beginning of each interval.

Similar to the person-interval approach, logistic regression or Cox regression models can be used to model the outcome of interest in each interval. However, as described in (Pepe et al., 1999), the main feature of this approach is that it allows model parameters to depend on both the timing of the desired prediction and the timing of the predictors:

$$\begin{aligned}logit\left\{Pr(Y_i(u) = 1|Y_i(t) = 0, W_i, Z_{i,t})\right\} &= \beta_0(u,t) + \beta_1(u,t)W_i + \beta_2(u,t)Z_{i,t} \\ \lambda_{i,t}(u|W_i, Z_{i,t}) &= \lambda_0(u,t)\exp\left(\beta_1(u,t)W_i + \beta_2(u,t)Z_{i,t}\right)\end{aligned}$$

Of note, the time-varying regression coefficients can be assumed to follow some smooth parametric functions, such as regression splines, which may involve both u and t. Furthermore, as intervals from the same individual are overlapping, correction for standard errors of the regression coefficients is required, preferably by using generalized estimating equation with an independence working correlation structure. As in the person-interval approach, the most general model in this setting is to fit totally different models at each time point t of measurement (Wagner et al., 1994; Rué et al., 2001). Further assumptions are required to simplify the model, for example, it is often sensible to only allow the intercept to depend on the time of the measurement (Lemeshow et al., 1994).

Dynamic prediction in this approach can be obtained in the same manner as in the person-interval approach. However, as the intervals are not restricted within a short predefined period, long-term prediction is possible for all models in this approach.

6.2.3 Approaches based on joint models

Even though the conditional modelling approaches described in the Sections 6.2.1 and 6.2.2 are easily interpretable and can be relatively easily implemented with standard statistical software, they rely only on observed values of the longitudinal process, which ignores the potential effect of measurement error, and they do not model the longitudinal data explicitly. Thus, it is not possible to formulate directly an association between the outcome and an underlying characteristic of the patient's entire covariate profile such as a constant slope of decline. Fortunately, these shortcomings can be resolved within the joint modelling framework. Joint models of a time-to-event or binary outcomes and longitudinal data have received a lot of attention in the statistical literature during the last years. The main purpose of this approach is to model the joint distribution of the outcome and the longitudinal data simultaneously (Verbeke and Davidian, 2008).

Essentially, joint modelling approaches use classical longitudinal models for the longitudinal data and logistic regression or survival analysis for the outcome but rather than being separate models, the two models are linked. In principle, the model for the outcome and the model for the longitudinal data can be linked in three different ways: (1) using observed values of the longitudinal process as covariates in the model for the outcome, (2) a two-stage approach in which first a longitudinal model is fitted (ignoring

the potential informative censoring induced by the outcome model) and then fitted values from that longitudinal model are used as covariates in the main outcome model, (3) using a shared latent structure for the two sub-models (Lawrence Gould et al., 2014). Even though the first two approaches are easy to implement with standard statistical software, they are somewhat ad-hoc and thus might produce biased results (Sweeting and Thompson, 2011). In contrast, the third approach specifies a proper probability model for the joint distribution of observed longitudinal and outcome data and is thus amenable to established statistical estimation methods such as maximum likelihood estimation. Therefore, most of the current research in this field focuses on this third approach, where the shared latent structure that links the main outcome model and the longitudinal processes model is either defined via shared random effects (shared random-effect models, SREM, (Wulfsohn and Tsiatis, 1997)) or via a latent class membership (joint latent class model, JLCM, (Proust-Lima and Taylor, 2009)).

In both SREM and JLCM approaches, the frequently used submodels for the longitudinal data and the main outcome, respectively, are the linear mixed effects model and the Cox proportional hazards model (for time-to-event outcomes) or logistic regression (for binary outcomes). Both approaches require the longitudinal and the outcome processes to be independent conditional on either the shared random effects (in SREM) or the latent class structure (in JLCM). JLCMs require a heterogeneous population of subjects that can be classified into multiple classes with different average longitudinal profiles and risks of outcome, while they do not rely on any specific assumptions regarding the relationship between the risk of the outcome and the longitudinal data in the model for the main outcome (Proust-Lima et al., 2014). In contrast, SREMs require assumptions regarding the effect of longitudinal data on the risk of the main outcome in the model for the main outcome. As the number of observed longitudinal measurements per individual decrease, parameter estimation in SREM becomes more sensitive to the assumptions regarding the distribution of the random effect (Rizopoulos et al., 2008); however, sparsity of longitudinal data might also a problem for the JLCM.

Parameters in both SREM and JLCM can be estimated using maximum likelihood estimation (Rizopoulos, 2012; Proust-Lima and Taylor, 2009). Methods for parameter estimation in shared random-effects joint models include approximate methods and likelihood-

based approaches using the EM algorithm (Tsiatis and Davidian, 2004). The more precise likelihood-based approaches are computationally intensive. Likelihood estimation for joint latent class models is more tractable because a closed-form of the (mixture) log-likehood can be derived and the parameters of interest can be estimated using standard maximum likelihood procedures (Proust-Lima and Taylor, 2009). Nevertheless, care must be taken as the likelihood function frequently has multiple maxima. Moreover, the model requires at least a verification of the latent class assumption. Parameters in SREM can also be estimated using a Bayesian approach (Faucett and Thomas, 1996). Within the Bayesian framework, computational implementation might be easier without the necessity of asymptotic approximations, and in situation where joint models are very complex and frequentist methods are infeasible, Bayesian approaches could provide a practical approach to solving the problem (Lawrence Gould et al., 2014).

A main advantage of joint models is that they model the joint distribution of longitudinal risk factor and outcomes efficiently and eliminate measurement error while providing valid inference. In both approaches, dynamic predictions can be obtained in the same way (Proust-Lima et al., 2014) and, in case of a survival model for the main outcome, the prediction is given by the following formulas:

$$\pi(u|t) = \sum_{g=1}^{G} Pr(T_i \leq u | T_i \geq t, c_i = g, W_i; \theta) Pr(c_i = g | T_i \geq t, Z_i(t), W_i; \theta)$$

$$\pi(u|t) = \int_{b_i} Pr(T_i \leq u | T_i \geq t, b_i, W_i; \theta) f(b_i | T_i \geq t, Z_i(t), W_i; \theta) db_i$$

where the first quantity in both formulas is the probability of outcome occurrence within the period (t, u) given the class membership (in JLCM, first formula) or the random effects (in SREM, second formula), and the second quantity is the probability that a subject belongs to a certain class (in JLCM) given current information or the density of random effects given current information (in SREM). More specifically, in the above formulas, T_i is the event time, c_i is the group membership (in a JLCM), b_i is the random effect (in a SREM), Z_i denotes longitudinal covariates, W_i denotes fixed covariates from both longitudinal and survival models, and θ denotes all model parameters (including regression coefficients).

In practice, one can calculate dynamic predictions from joint models by plugging-in pa-

rameter estimates and empirical Bayes estimates for random effects into an approximated version of the above formula. Corresponding standard errors and confidence intervals for predictions can be obtained by approximating the distribution of dynamic predictions using Monte Carlo simulation (Rizopoulos, 2012).

Joint models have been successfully applied to address prognostic questions in several clinical settings including the prediction of relapse of prostate cancer based on longitudinal post-treatment PSA measurements (Proust-Lima and Taylor, 2009) and the prediction of rupture of the abdominal aortic aneurysm (AAA) based on AAA diameter measurement via ultrasound (Sweeting and Thompson, 2011). However, the resulting models and predictions are difficult to interpret for a clinical audience. Moreover, these models appear to be most suitable for rich datasets with extensive longitudinal data and one (or a low number) of different longitudinal markers only. Current research includes novel suggestions to develop joint models with more than one longitudinal marker (Andrinopoulou et al., 2014; Rizopoulos and Ghosh, 2011).

6.3 Assessment of dynamic prognostic models

In general, all performance criteria described in Section 2.2.3 can be used to assess dynamic prognostic models. Amongst them, the Brier score (for overall performance) and the AUC (for discrimination) are the most frequently used criteria (Schoop et al., 2008, 2011; Zheng and Heagerty, 2007; Rizopoulos, 2011; Blanche et al., 2014). For prediction models using baseline information only, model performance only depends on the prediction window. However, in the dynamic prediction framework, model performance depends on both the time point of the prediction and the prediction window. Therefore, plots which describe changes in performance of each model depending on either the prediction time point for a specific prediction window, or depending on the prediction window at a specific prediction time point can be used to compare performance between models (Proust-Lima and Taylor, 2009). In addition, as updated predictions are only relevant to observations still at risk, performance assessment at each prediction time point in the dynamic prediction scheme is restricted to the at-risk population at that time (Schoop et al., 2008).

When the outcome of interest is the time to an event of interest, both baseline and dynamic prediction frameworks have to take into account censoring, as contributions to the Brier score or AUC from individuals who are censored before the future time of interest u cannot be defined. When censoring is assumed to be independent of the time to event and the longitudinal processes, the inverse probability of censoring weighting technique (Graf et al., 1999; Blanche et al., 2014) can be used to make the population of non-censored individuals (up to a future time u) representative of the whole at-risk population at prediction time t by up-weighting their contributions to the performance measure with weights defined as

$$w_i(u,t) = \frac{I(T_i > u)}{G(u|t)} + \frac{I(t < T_i \leq u)\Delta_i}{G(T_i|t)}$$

where T_i is the individual observed follow-up time, δ_i is the individual event indicator (1 if event occur, 0 if being censored), $G(u|t)$ is the probability of not being censored at time u given not being censored at time t, $G(T_i|t)$ is the individual probability of not being censored at the end of follow-up time given not being censored at time t.

As in the traditional framework, developing a prediction model on a large dataset and then validating it on an external dataset is also recommended for dynamic prediction models (Proust-Lima et al., 2014). When external validation is impossible, internal validation using e.g. cross-validation can be used to correct for optimism; however, this strategy is computational intensive for complex dynamic models such as joint models.

6.4 Differences between acute and chronic disease settings (and implications for modelling)

Chronic diseases are diseases with a long duration and slow progression, such as cancer, cardiovascular or liver diseases. In contrast, acute diseases progress rapidly within a short duration, as is the case for many infectious diseases and in emergency care. Regarding dynamic prediction modelling, it is interesting to note that complex approaches (joint models) have been developed and applied mainly for chronic diseases (Proust-Lima and Taylor, 2009; Rizopoulos, 2012; Sweeting and Thompson, 2011), whereas, simpler ap-

proaches (conditional models) have been applied in both acute (Ruttimann and Pollack, 1991; Rué et al., 2001; Lemeshow et al., 1988, 1994; Wagner et al., 1994) and chronic settings (Christensen et al., 1993; Hughes et al., 1992; Wu and Ware, 1979; Hartmann et al., 2012; Cupples et al., 1988; Murtaugh et al., 1994; Van Houwelingen and Putter, 2012). This divergence in the application of dynamic models suggests that there are differences between the chronic and the acute disease settings, which may affect the development and application of dynamic prediction models.

6.4.1 Time origin, prediction horizon, and outcome of interest

One important feature of acute diseases is that the disease (especially in infectious diseases) often has a clear time origin (time of infection, for example) and only lasts for a certain period. After that time, the disease usually resolves and the patient fully recovers. Therefore, the prediction horizon in this setting is restricted to a specific period where the event of interest may occur, and predictions beyond that period are of no interest. As time evolves, the clinically useful time horizon of predictions decreases, and in some sense, long-term predictions converge to short-term predictions. Of note, early prediction is key in acute diseases as predictions at a late time point allow only for a very limited remaining time window for possible interventions. Therefore, in acute disease settings, the prediction time is restricted to a specific period in the early phase of the disease in order to be clinically useful.

In contrast, chronic diseases usually have no clear time origin as the disease can progress slowly while in a "hidden" state before becoming clinically apparent (Liestø l and Andersen, 2002). Therefore, the time origin in this setting is usually defined as the time of diagnosis, the start date of an intervention or even the somewhat arbitrary time point when the patient was enrolled into a prognostic study. Furthermore, the definition of "cure" is vague in this setting and usually refers to some arbitrary fixed time interval, for example, "recurrence-free survival for >5 year" (Van Houwelingen and Putter, 2012). Thus predictions at any time point of the disease for a fixed prediction horizon may be clinically useful in this setting accepting that the patient will still suffer from the chronic disease at the end of the prediction period.

In addition, as "cure" is clearly defined in most acute diseases and many patients are

hospitalized for treatment or at least under close observation as outpatients during the entire disease period, censoring is often not a major issue in acute diseases. Also, the observed follow-up duration is usually the same for all subjects. In contrast, in chronic diseases which require long-term follow-up of patients, losses to follow-up are an important problem and the validity of statistical models may depend heavily on the amount and mechanism of censoring (Fitzmaurice et al., 2004).

These differences between the two settings might also affect the decision regarding the outcome of interest, and the statistical model for that outcome. For example, in acute diseases, the outcome is often a binary indicator of the occurrence of a disease event of interest at any time point during the relatively short disease course. Hence, a logistic model is the model of choice. In contrast, in chronic diseases with longer and often unequal follow-up of patients, the time to an outcome might be more relevant and a survival regression model could be a reasonable model.

6.4.2 Repeated measurement

As many acute diseases require hospitalization for monitoring and treatment, longitudinal information is usually recorded regularly with a common schedule for all patients, resulting in balanced and complete longitudinal data. However, for the chronic setting, longitudinal data is more irregularly collected and individual patients may delay or miss scheduled follow-up visits. On the other hand, as the course of disease is relatively short, the number of repeated measurements per patient is often limited in the acute setting.

These differences may affect how longitudinal data are modelled. For example in the person-interval approaches described in Section 6.2.2, models developed in the chronic setting may depend on how each person-interval is defined; whereas this may not be an important issue in the acute setting as the longitudinal dataset is balanced and thus different interval splitting strategies lead to the same result. One the other hand, limitations regarding the number of repeated measurement per patients may restrict the use of complex and flexible models such as joint models in the acute setting.

6.4.3 Relationship between outcome and time-dependent covariates

In acute diseases, the whole time course of disease can often be virtually divided into different latent phases: onset, critical and recovery phases, and outcomes of patients may be very different between these phases. As a result, the assumption of a time-independent relationship between the outcome and time-dependent covariates may not hold in the acute setting, but in chronic diseases which are more stable over time, the assumption may be valid. Furthermore, in acute diseases the current value of a biomarker, which may reflect the current response of a patient to treatment, could be the most relevant predictor for outcome from that time point onwards. However, as disease progress is slow in chronic diseases, it is usually reasonable to argue that the whole trajectory of repeated measurement is required to predict outcome accurately. In practice, these differences might affect how the relationship between outcome and time-dependent covariates is specified in dynamic prediction models.

6.4.4 Competing risks

As such illnesses evolve over a long time span, progression of a specific chronic disease can be complicated by the presence of other diseases. Therefore, many types of event may occur and some of them may affect the occurrence of the main event of interest. This "competing risks" problem would require special consideration in the modelling steps in order to provide valid predictions (Wolbers et al., 2009). However, in acute disease settings, competing risks are rare and often biologically implausible; therefore, this issue can be ignored when developing prediction models in this setting.

6.4.5 Clinical usefulness

As acute diseases often require prompt management decisions within a short time, a prognostic model must be easy to interpret and easy to use, in order to be widely used in clinical practice. From this practical point of view, complex models such as joint models might be inferior to simpler models, as it is difficult not only to explain them to a non-statistical audience but also to retrieve outcome predictions. This drawback may hamper the implementation of such models in the field of acute diseases.

6.5 Case study: dynamic prediction models for the development of DSS in hospitalized dengue patients

In chapter 5, a prediction model for the development of DSS with 7 covariates including age, gender, day of illness, history of vomiting, temperature, having a palpable liver, and platelet count was derived using data from 2301 children hospitalized with dengue infection. However, this model only had a moderate performance in both temporal and internal validation. Of note, platelet count, a well-known risk factor of DSS, was recorded daily in that study. Therefore, there is an opportunity to assess whether integrating this longitudinal information can improve the performance of the baseline model presented in the previous chapter.

6.5.1 Description of data

For the purpose of this cases study, only patients who enrolled into the MD cohort on day 3 of illness were included. The main reason for this is that including all patients would both complicate statistical modelling and clinical interpretation of a dynamic prediction model. The previous baseline model involves several clinical signs and symptoms, which are time-dependent. However, the information regarding these covariates was only collected at the single time point of enrolment. Including only individuals who enrolled on the same day of illness unifies the time scale from disease onset (which is the most clinically relevant scale) and the time scale from enrolment. Hence signs and symptoms can simply be regarded as baseline covariates and there is no need to model time-varying signs and symptoms and their effect on outcome, which would require strong and untestable assumptions as only a single measurement per patient is available. In addition, even though there were more patients enrolled on day 4 than day 3 of illness, the time point at day 3 was still chosen, as it is more useful to obtain prediction of DSS development from day 3 onwards rather than from day 4 which is too close to the time that DSS often occurs.

Amongst all 908 confirmed dengue patients from the MD cohort who enrolled on day 3, 17 patients did not have a platelet count at enrolment and were excluded from the analysis. Therefore, the final analysis included data at enrolment and updated platelet counts from 891 patients.

Clinical outcome

In total, 59 cases (59/891, 7%) developed DSS (Table 6.1). While DSS can occur at any time from day 4 to day 8 of illness, most patients developed DSS within the first 2 days after enrolment (day 4-5 of illness). For patients who did not develop DSS, a few were discharged early but 95% remained in the hospital until illness day 7 or later (Table 6.1).

There are two clinical outcomes of interest in this case study: (1) whether a patient progresses to DSS at all, and (2) whether a patient progresses to DSS on the following day given the present state of the patient. The former question refers to long-term prediction, while the latter refers to short-term prediction. The course of dengue infection only lasts for 1-2 weeks and in the dataset only a single DSS case occurred after day 7 of illness. Based on this I chose days 3-6 of illness as the relevant prediction time points. Furthermore, it is reasonable in this setting to assume that patients will not develop DSS after hospital discharge. Therefore, rather than using the real follow-up time of each patient, which would imply that the patient's disease status after discharge is unknown, I reset the follow-up times for all patients discharged without DSS to 6 days after enrolment (i.e. day 9 of illness). Let T_i be the day of illness on which DSS occurred (which was on day 8 or earlier for all subjects) for subjects with DSS, and day 9 for subjects without DSS. The long-term outcome can be rephrased in terms of follow-up time as $(T_i \leq 8)$; therefore, the long-term prediction made at time t (with $t \leq 6$) is $Pr(T_i \leq 8|T_i > t)$. Similarly, the short-term prediction at time t is $Pr(T_i = t+1|T_i > t)$ for $t \leq 6$.

Longitudinal data

Longitudinal data in this study includes daily platelet counts of each patient until DSS development or discharge. The majority of subjects had 4-6 measurements but for patients who developed DSS, most of them only had 1-3 measurements (Table 6.1). Amongst all patients, 35 (4%) cases had at least one missing platelet count within their series (31/35 had only 1 missing value, 4/35 cases had 2 missing values).

Table 6.1. Outcome and number of platelet counts per patient in this case study (n = 891).

Characteristics	N	(%)
DSS	59	(7)
Day of DSS		
- Day 4 of illness	22	(37)
- Day 5 of illness	21	(36)
- Day 6 of illness	10	(17)
- Day 7 of illness	5	(8)
- Day 8 of illness	1	(2)
Day of discharge (in patients without DSS)		
- Day 4 – 6 of illness	41	(5)
- Day 7 – 8 of illness	428	(51)
- Day 9 of illness or later	363	(44)
Number of platelet count measurements per patient		
- 1 to 3	75	(8)
- 4 to 6	704	(79)
- 7 to 9	112	(12)
Number of platelet count measurements before DSS (in patients with DSS)		
- 1	22	(37)
- 2	21	(36)
- 3	11	(19)
- 4	4	(7)
- 5	1	(1)

Other covariates

In this analysis, the baseline variables that were identified as risk factors of DSS according to the analysis of Chapter 5 were also included: age, gender, history of vomiting, temperature, and having a palpable liver. There were 14/891 (1.6%) cases with at least one missing value for these covariates (1 for temperature, 11 for liver size, 2 for history of vomiting). As the number of missing values was low, I chose to impute these missing values by using the category with the highest frequency for categorical variables and the median of observed values for continuous variables.

6.5.2 Exploratory analysis of repeated platelet counts and their potential benefit for the prediction of DSS development

Figure 6.2 describes the trajectories of the platelet count over the course of disease for all patients included in this analysis. On average, the platelet count tended to decrease

initially until about day 6, and then slowly returned to the normal level when the disease resolved.

Interestingly, platelet counts in patients who developed DSS tended to be lower than in patients who did not have DSS, and this difference was most pronounced on the day before DSS occurrence (Figure 6.3). This observation suggests that the platelet value on the current day or the change from the previous day may relate to the occurrence of DSS on the next day.

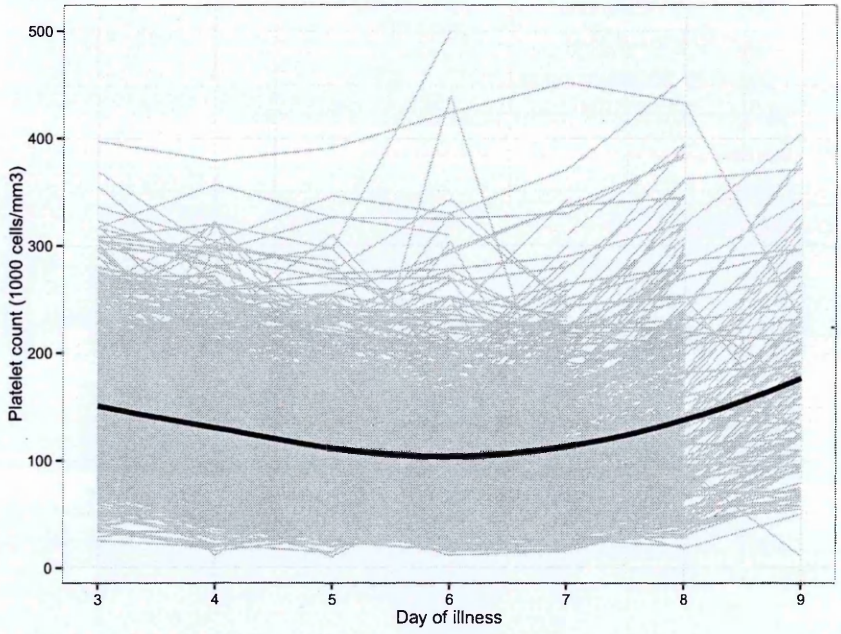

Figure 6.2. *Individual trajectories of platelet counts from day 3 to day 9 of illness amongst all 891 patients in this analysis (grey lines). The black line displays a loess scatterplot smoother.*

To investigate the relationship between the risk of DSS development and the platelet count on a specific day of illness further, patients still at risk on that day (i.e. those without DSS until that day) were split into groups of equal size, based on their platelet values (current or previous values, or current change). Then, the average platelet count and average change was calculated in each group and compared to the percentage of subjects who developed DSS on the next day or overall in that group, respectively. These values are displayed in Figures 6.4 and 6.5, which show a negative relationship between the current platelet count and the change in the platelet count from the previous day with both the short-term and long-term occurrence of DSS. Based on these observations, the current platelet count or the change from the previous day could be relevant to predicting

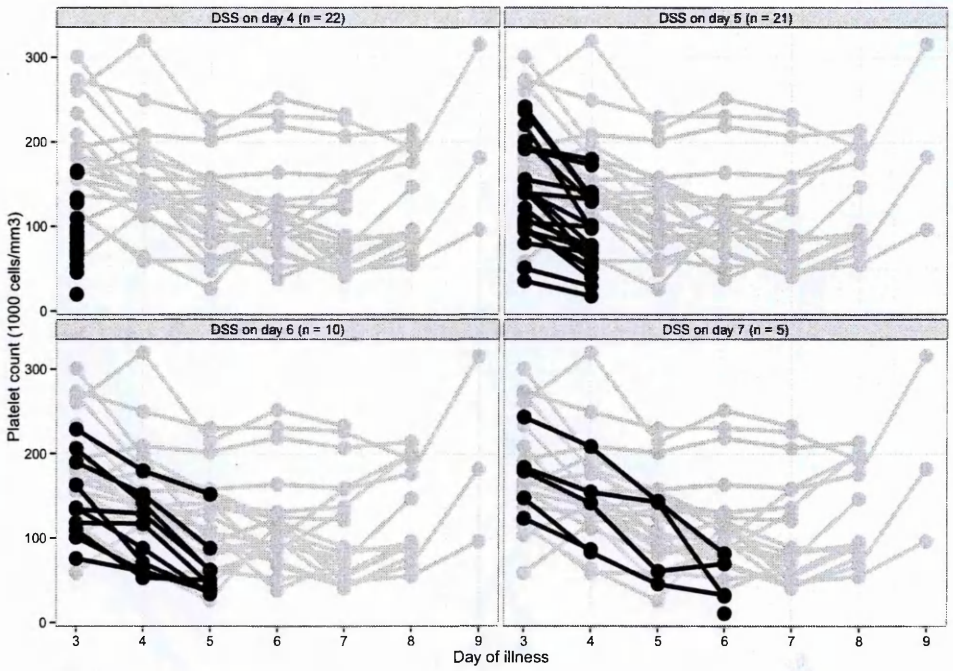

Figure 6.3. *Trajectories of platelet counts for all patients who developed DSS from day 4 to day 7 of illness (black lines and dots) and 20 randomly chosen patients who did not have DSS (grey lines and dots).*

the occurrence of DSS over time. As the change in platelet count cannot be determined at the time of enrolment, the current platelet count was the main variable of interest in the development of dynamic prediction models in the following section.

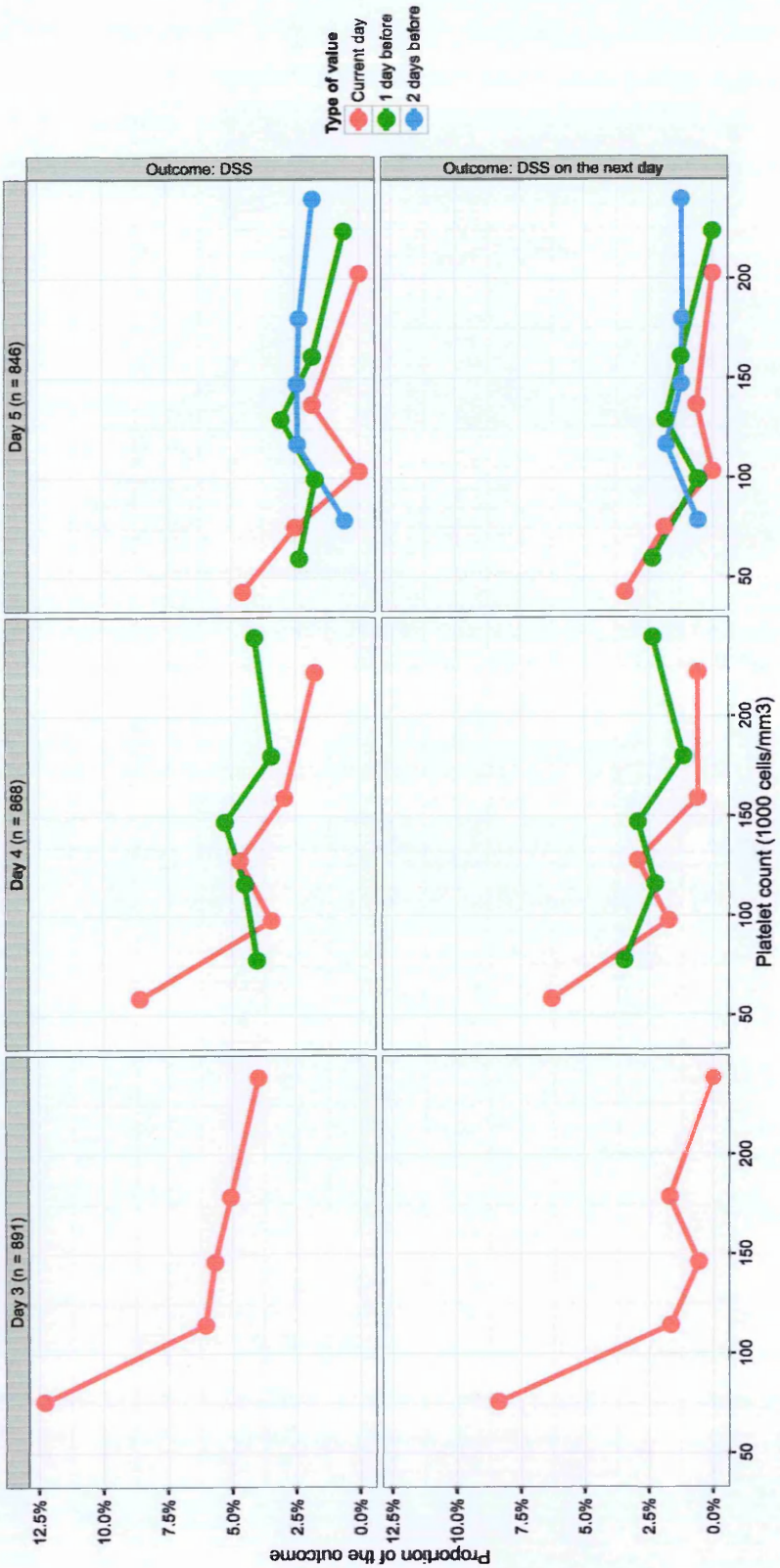

Figure 6.4. *Relationship between platelet counts at different time points and the risk of long-term outcome (overall DSS occurrence) or short-term outcome (DSS occurrence on the next day) amongst patients still at risk on days 3, 4, and 5 of illness. For each population at risk, individuals were grouped into 5 groups of equal size based on their current platelet value (or values from the previous 2 days, respectively). Dots display the mean platelet count and the observed risk of the outcome in each patient group. n refers to the total sample size of each at-risk population.*

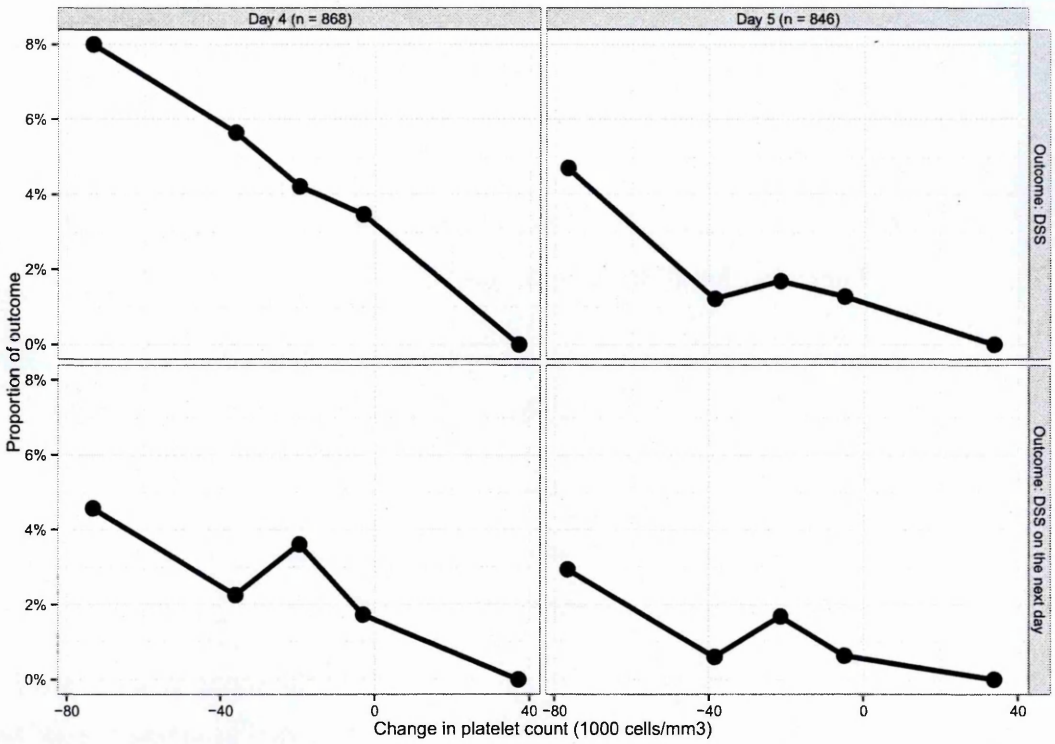

Figure 6.5. *Relationship between changes in platelet count from the previous day (value on the current day minus value on the previous day) and the risk of long-term outcome (overall DSS occurrence) or short-term outcome (DSS occurrence on the following day) amongst patients still at risk on day 4 and 5 of illness. For each population at risk, individuals were grouped into 5 groups of equal size based on the observed changes in their platelet counts. The dots display the mean change and the observed risk of the outcome in each patient group. n refers to the total sample size of each at-risk population.*

6.5.3 Dynamic prediction modelling – model specification and assessment

Model specification

I compared a baseline model, which included only the platelet count at enrolment, to several dynamic prediction models for the risk of a short or long-term outcome (DSS occurrence on the next day or overall DSS occurrence, respectively) based on the approaches to model development described in Section 6.2. The candidate dynamic prediction models covered a model using the person-interval approach (for short-term outcome only), two partly conditional models which included either the current platelet value alone or the current value and the change from the previous value, and a joint model. A detailed specification of the models is provided below.

The specification of the models is based on the following notation:

- t denotes the day of illness (predictions at time $t = 3, 4, 5, 6$ are desired to predict the short-term outcome at time $t + 1$ and the long-term outcome after time t).
- T_i denotes the day of illness on which DSS occurred in patient i (which was on day 8 or earlier for all subjects) and day 9 for subjects without DSS.
- W_i is the vector of time-fixed covariates for patient i (age, gender, history of vomiting, temperature, and having a palpable liver).
- $Z_i(t)$ is the observed platelet count at day t. The observed count at enrolment is denoted by both $Z_i(3)$ to emphasize its usage as a time-dependent value and $Z_{i,3}$ to emphasize its usage as a baseline covariate.
- $\Delta_i(t) = Z_i(t) - Z_i(t-1)$ is the change in platelet count from the previous day. The change is only defined from illness day 4 onwards as values before day 3 are not available.
- The discrete time hazard on day t is denoted by $\lambda(t)$ and defined as $\lambda(t) = Pr(T_i = t|T_i > t-1)$. $\lambda_0(t)$ refers to the baseline hazard.
- All regression coefficients are denoted by β with corresponding subscripts. Regression coefficients that depend on time are denoted by $\beta(t)$.

Models for short-term prediction (DSS on the next day) As the baseline model, a traditional discrete time Cox proportional hazards regression model was used which depended only on baseline information:

$$Pr(T_i = t+1|T_i > t, W_i, Z_{i,3}) = 1 - \exp\{-\exp(\beta_0(t) + \beta_1 W_i + \beta_2 Z_{i,3})\}$$

This is denoted as the *baseline Cox model* in the following paragraphs.

The *person-interval model* split the follow-up time of each patient into distinct intervals of length one day and applied a binary regression model with a time-varying intercept to the pooled data from all intervals:

$$g\{Pr(T_i = t+1|T_i > t, W_i, Z_i(t))\} = \beta_0(t) + \beta_1 W_i + \beta_2 Z_i(t)$$

$g()$ in the formula above denotes the link function for which I chose the complementary log-log link. This implies that the fitted model is equivalent to a discrete time Cox proportional hazards model with a time-dependent covariate (Singer and Willett, 2003).

The *partly conditional model* splits the follow-up time of each patient into overlapping intervals starting from the prediction time point t_0 to the maximum follow-up of day 9. The following Cox regression model was fitted to the pooled data set of information from these intervals

$$Pr(T_i = t+1|T_i > t \geq t_0, W_i, Z_i(t_0)) = 1 - \exp\Big\{-\exp\Big(\beta_0^{(t_0)}(t) + \beta_1 W_i + \beta_2 Z_i(t_0)\Big)\Big\}$$

This is a discrete time Cox model with time-independent covariates stratified by the prediction time point t_0 (i.e. allowing for separate baseline hazards for each t_0). This model is referred to as *"partly conditional survival model (1)"*.

A second partly conditional model (*"partly conditional survival model (2)"*) was also investigated to assess whether adding the change in the platelet count as a covariate improves prediction. Specifically, this model has the following form:

$$\begin{aligned} &Pr(T_i = t+1|T_i > t \geq t_0, W_i, Z_i(t_0), \Delta_i(t_0)) \\ &\qquad = 1 - \exp\Big\{-\exp\Big(\beta_0^{(t_0)}(t) + \beta_1 W_i + \beta_2 Z_i(t_0) + \beta_3 \Delta_i(t_0)\Big)\Big\} \end{aligned}$$

As the change is only available from day 4 onwards, this model was only fitted to prediction time points t_0 with $t_0 \geq 4$.

Finally, a *joint model* was fitted. Based on Figure 6.2, the trajectory of platelet count for each patient could be modelled by a linear mixed effects model with a quadratic function for platelet count over time and allowing for individual variation by using random intercept and slope terms. Moreover, platelet count is known to depend on gender and age. Hence, the longitudinal sub-model of the joint model was defined as follows:

$$Platelet_i(t) = Z_i^*(t) + \varepsilon_i(t)$$
$$= \alpha_0 + \alpha_{0,1}Age_i + \alpha_{0,2}Gender_i + \alpha_1 t + \alpha_2 t^2 + a_{0,i} + a_{1,i}t + \varepsilon_i(t)$$

Here, $Z_i^*(t)$ denotes the "true" platelet value without measurement error $\varepsilon_i(t)$ and the random effects $a_{0,i}$ and $a_{1,i}$ are assumed to have a joint bivariate normal distribution independent of measurement error $\varepsilon_i(t)$. The survival sub-model of the joint model is a Cox regression model that included the true current platelet count (without measurement error) as a covariate

$$Pr(T_i = t+1 | T_i > t, W_i, Z_i^*(t)) = \lambda_0(t+1)\exp(\beta_1 W_i + \beta_2 Z_i^*(t))$$

In this model, the log baseline hazard function was modelled using regression splines with knot locations chosen automatically by the statistical software (Rizopoulos, 2010).

Models for long-term prediction (DSS on any subsequent day) For the model including baseline information only, I chose a binary regression model (the *baseline binary model*) with an intercept that varies by day of illness:

$$g\{Pr(T_i \leq 8 | T_i > t, W_i, Z_{i,3})\} = \beta_0(t) + \beta_1 W_i + \beta_2 Z_{i,3}$$

For this model and all subsequent models, $g()$ denotes the complementary log-log link for consistency reasons with the partly conditional model for short-term survival described above. Of note, person-interval models are not designed for long-term prediction and hence were not implemented for this purpose.

The *partly conditional binary models* split the follow-up time of each patient into overlapping intervals starting from the prediction time points t to the maximum follow-up of day 9. They then applied a binary regression model to the pooled data including the current platelet count (*partly conditional binary model (1)*) or the current count and the change from the previous day (*partly conditional binary model (2)*, only for illness day 4 onwards) leading to the following models:

$$g\{Pr(T_i \leq 8|T > t, W_i, Z_i(t))\} = \beta_0(t) + \beta_1 W_i + \beta_2 Z_i(t)$$

and

$$g\{Pr(T_i \leq 8|T > t, W_i, Z_i(t))\} = \beta_0(t) + \beta_1 W_i + \beta_2 Z_i(t) + \beta_3 \Delta_i(t)$$

Finally, the *joint model* described above can also be used for long-term prediction.

Model estimation

As prediction times of interest were day 3, 4, 5, and 6, only data of repeated measurement up to day 6 of illness were used to estimate parameters in all models. When the observed platelet count at prediction time t was missing, its value was substituted by the last non-missing observed value up to that time point (the "last observation carried forward" imputation method (Collett, 2003)). All models were estimated within the likelihood framework.

In this setting, time is discrete leading to many ties, i.e. DSS events recorded on the same day. Therefore, instead of using standard continuous-time survival model, I used discrete-time approaches for all survival models throughout, except for the survival submodel of the joint model (as this feature has not yet been implemented in current statistical software). Similar to the person-interval approach of dynamic prediction modelling, fitting a discrete-time survival model requires splitting the follow-up time of each subject into short time intervals and then using the pooled dataset for model estimation. The statistical model of choice in this case is binary regression model with a complementary log-log link which can be estimated using standard software for generalized linear models (Singer and Willett, 2003). Hence, the main function for fitting the models was the function glm() which fits generalized linear models in the statistical software R (R Core Team, 2014).

Generalized estimating equations with an independence working correlation structure, as implemented in package geepack (Halekoh and Højsgaard, 2006), were used to correct the estimated standard errors in the baseline binary regression model and partly conditional models. Finally, the joint model was fitted using package JM (Rizopoulos, 2010).

Prediction and model assessment

Short-term and long-term predictions are straightforward to obtain from the respective short- and long-term prediction models other than the joint model. For the joint model, both short-term and long-term predictions were obtained by plugging-in parameter estimates and empirical Bayes estimates for random effects into an approximated function of the dynamic predictive distribution, as described in Section 6.2.3 and implemented in the aforementioned R package JM. If a patient's platelet count at a prediction time point was missing, then this patient was excluded from the respective data set for predictions.

Between days 3-6 of illness the performance of short- and long-term predictions of all models amongst patients at risk (i.e. those without DSS at or before the prediction time point) was evaluated with the Brier score and the area under the ROC curve (AUC). As no independent data is available for external validation, internal validation on the original dataset with 10-fold cross validation was used to correct for optimism.

6.5.4 Dynamic prediction modelling – results

Tables 6.2 and 6.3 summarize the estimated regression coefficients for each developed model. The effects of time-fixed covariates on outcomes were similar between models. However, the effects of the baseline platelet count on outcomes were smaller than the corresponding effects of the current platelet count. The results also showed a bigger effect of the current platelet count on short-term outcome in the joint model compared to person-interval and partly conditional survival models, which can be explained by the attenuation effect of measurement error.

Table 6.2. Estimated coefficients (Est) and corresponding standard errors (SE) from fitted models for short-term prediction of DSS.

Covariate	Baseline Cox model		Person-interval model		Partly conditional survival model (1)		Partly conditional survival model (2)		Joint model	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Age [+ 1 year]	-0.05	0.06	-0.08	0.06	-0.04	0.05	-0.01	0.06	-0.15	0.07
Gender: Females	-0.30	0.28	-0.43	0.28	-0.47	0.32	-0.57	0.43	-0.50	0.30
History of vomiting: Yes	1.00	0.27	1.05	0.27	0.91	0.33	0.79	0.42	0.94	0.29
Temperature [+1°C]	0.44	0.19	0.46	0.20	0.44	0.18	0.42	0.20	0.47	0.22
Palpable liver: Yes	0.39	0.37	0.10	0.39	0.47	0.46	0.63	0.58	0.01	0.41
PLT (baseline) [+10,000 cells/mm3]	-0.09	0.03								
PLT (current) [+10,000 cells/mm3]			-0.23	0.03	-0.11	0.03	-0.11	0.04	-0.45	0.08
PLT (change) [+10,000 cells/mm3]							-0.09	0.02		

The baseline Cox model included the baseline platelet count only, whereas all other models used the current platelet count as a covariate. The partly conditional survival model (2) additionally included the change in the platelet count from the previous day and was fitted to data from 4 onwards only. For the joint model, the regression coefficient for the current platelet value refers to the effect of the unobserved measurement-error free value.

Table 6.3. *Estimated coefficients (Est) and corresponding standard errors (SE) from fitted models for long-term prediction of DSS.*

Covariate	Baseline binary model		Partly conditional binary model (1)		Partly conditional binary model (2)	
	Est	SE	Est	SE	Est	SE
Age [+ 1 year]	-0.02	0.05	-0.04	0.05	-0.01	0.07
Gender: Females	-0.38	0.32	-0.47	0.32	-0.57	0.42
History of vomiting: Yes	0.85	0.33	0.90	0.33	0.77	0.42
Temperature [+1°C]	0.43	0.18	0.44	0.18	0.43	0.21
Palpable liver: Yes	0.63	0.49	0.46	0.48	0.64	0.60
PLT (baseline) [+10,000 cells/mm^3]	-0.05	0.03				
PLT (current) [+10,000 cells/mm^3]			-0.11	0.03	-0.11	0.04
PLT (change) [+10,000 cells/mm^3]					-0.09	0.03

The baseline model included the baseline platelet count only with a time-varying intercept whereas all other models used the current platelet count as a covariate. The partly conditional binary model (2) additionally included the change in the platelet count from the previous day and was fitted to data from day 4 onwards only.

The Brier score and the AUC at different prediction time points are displayed for all models in Figures 6.6 and 6.7 (short-term prediction) and Figures 6.8 and 6.9 (long-term prediction). In terms of overall performance, the Brier score revealed no apparent differences between models. In terms of discrimination, AUCs were higher for short-term prediction compared to long-term prediction. The baseline models were inferior to all other models for both short-term and long-term prediction. Furthermore, AUCs of the baseline models for short-term prediction tended to decline as the prediction time point increased whereas the performance of other models was relatively stable over time, which suggests a decrease in the relevance of the day 3 platelet count on short-term prognosis at later time points. Differences in discrimination between other models, which required updated platelet counts, were minimal. Adding the change in platelet count from the previous day to the model in addition to the current platelet count seemed to increase discrimination of these models; however, the improvement was not remarkable. Of note, assessing performance of a single model over time based on long-term predictions might be misleading as the prediction horizon differs for different prediction time points.

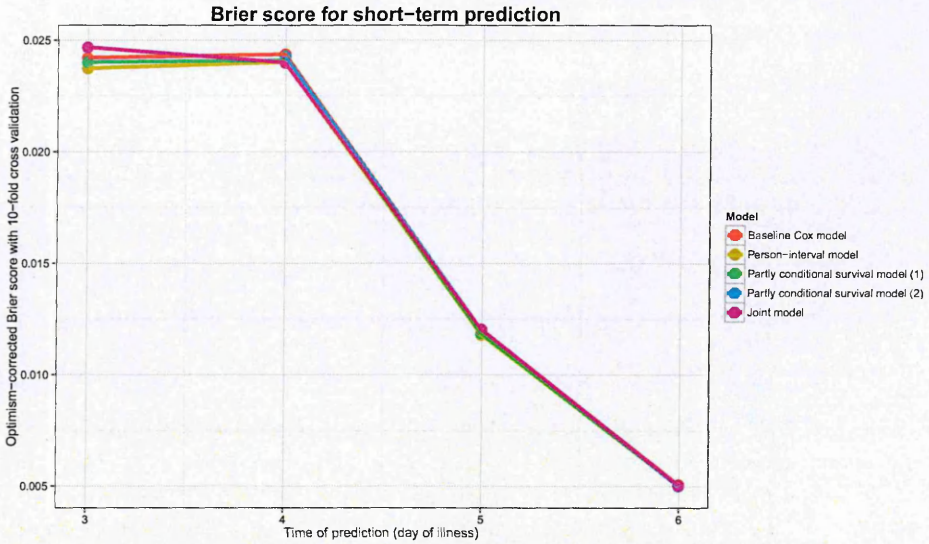

Figure 6.6. *Brier score for short-term prediction (probability of having DSS on the next day) of each model at each prediction time. All values were corrected for optimism via 10-fold cross validation. Partly conditional survival model (1) only used the current platelet count whereas partly conditional survival model (2) included both the current value and the change from the previous value as covariates. Predictions for the partly conditional survival model (2) are only available from illness day 4 onwards.*

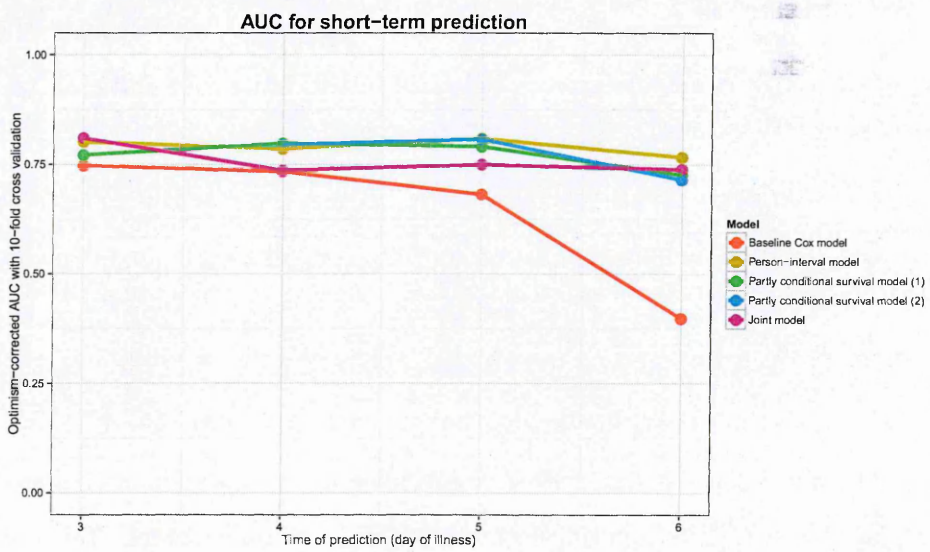

Figure 6.7. *Area under the ROC curve (AUC) for short-term prediction (probability of having DSS on the next day) of each model at each prediction time. All values were corrected for optimism via 10-fold cross validation. Partly conditional survival model (1) only used current platelet count while partly conditional survival model (2) included the current value of the platelet count and the change in value from the previous day as covariates. Predictions for the partly conditional survival model (2) are only available from illness day 4 onwards.*

6.6 Discussion

To date, the majority of dynamic prediction models have been developed for chronic diseases. In this chapter, I contrasted several aspects of chronic versus acute diseases which affect the choice of the most appropriate statistical model. The period during which an acute disease evolves is usually much shorter and patients are often under close observation throughout the disease. As a consequence, the resulting dataset is more likely to contain balanced longitudinal data collected at the same discrete time points for each patient, to include the same duration of follow-up for each patient which is sufficient to conclusively assess the outcome of the disease, and to have little missing data. In this sense, the development of dynamic prediction models in acute diseases might be considered to be easier than in chronic diseases. However, due to the limitations in the amount of available longitudinal data and the dynamic nature of the disease, developing sophisticated and flexible models such as joint models is not always possible in acute diseases, and if it is, there is no guarantee that these models would perform better than simpler ones.

Based on the presented case study of dynamic prediction modelling in hospitalized dengue patients where DSS was the outcome of interest, several observations can be made. First, as longitudinal data was balanced, it was quite easy and useful to explore the potential value of longitudinal information by stratifying patients based on the prediction time and using simple graphical tools. Second, the case study clearly demonstrated the usefulness of dynamic prediction modelling as all investigated dynamic models outperformed the models that included baseline information only. This is in accordance with similar findings for many other diseases (Lemeshow et al., 1988; Christensen et al., 1993; Hughes et al., 1992; Rué et al., 2001; Karp et al., 2004). Third, all the dynamic models investigated had a similar performance and the simpler conditional models even tended to have a slightly superior performance than the joint model. This suggests that in the acute setting where longitudinal data is often balanced but also limited, simple approaches (conditional models) are indeed preferred to complex models. Of note, I included only one longitudinal covariate in the case study as changes in symptoms over time were not recorded in the MD study. However, extensions to more than one longitudinal covariate

are straightforward in the conditional setting whereas joint modelling is challenging for multivariate longitudinal data and publicly available software implementations to fit such models are currently lacking.

Specific to dengue infection, this case study provides evidence that the current platelet value or the change in value from the previous day could be used to improve prediction of the occurrence of DSS, especially for short-term prediction. While reliable long-term predictions would be desirable, short-term predictions could still be useful in supporting the day-to-day management of patients, for example the decision whether daily outpatient follow-up is sufficient for a patient or they require hospitalization. A limitation of the present dataset is that the sample size and the number of DSS cases was too small to draw definite conclusions. This also prevented the exploration of more complex models with time-varying coefficients for the longitudinal platelet count or non-linear platelet effects.

In conclusion, this chapter suggests that dynamic prediction models based on conditional models, which are relatively straightforward to implement, can improve prediction in acute diseases where longitudinal data is frequently routinely collected. In dengue, a large international cohort study which collects detailed longitudinal laboratory data as well as signs and symptoms of dengue patients is currently recruiting (Jaenisch et al., 2013). The resulting data set will open up an opportunity for dynamic prediction modelling which could lead to improved case management and the early identification of cases likely to develop DSS.

6.7 Appendix

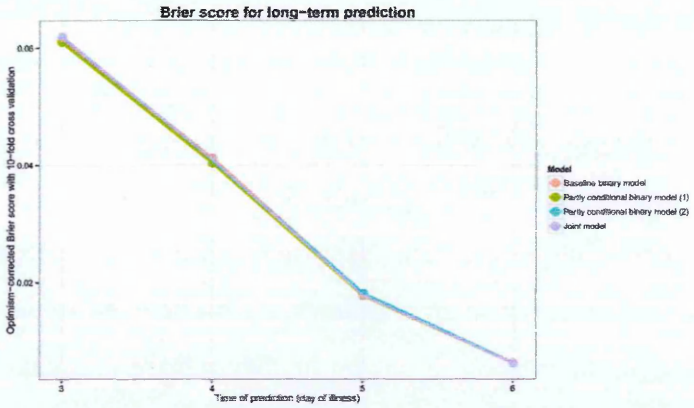

Figure 6.8. *Brier score for long-term prediction (overall DSS occurrence) of each model at each prediction time. All values were corrected for optimism via 10-fold cross validation. Partly conditional binary model (1) only used the current platelet count whereas partly conditional binary model (2) included both the current value and the change from the previous value as covariates. Predictions for the partly conditional binary model (2) are only available from illness day 4 onwards.*

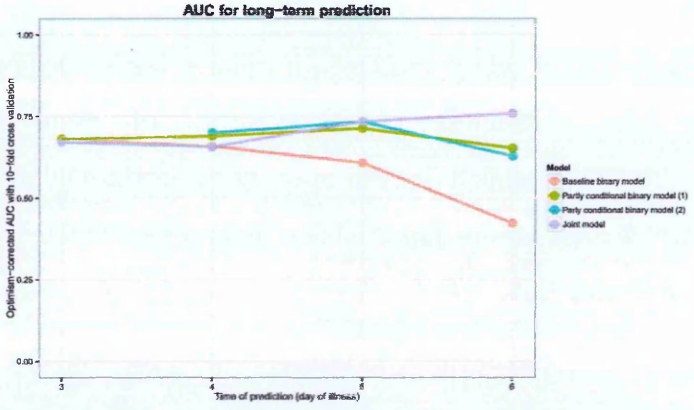

Figure 6.9. *Area under the ROC curve (AUC) for long-term prediction (overall DSS occurrence) of each model at each prediction time. All values were corrected for optimism via 10-fold cross validation. Partly conditional binary model (1) only used current platelet count while partly conditional binary model (2) included both current value and change in value from previous day of platelet count as covariates. Predictions for the partly conditional binary model (2) are only available from illness day 4 onwards.*

Chapter 7

Conclusions

7.1 Contributions of this thesis

7.1.1 Clinical contributions

This thesis provides the first comprehensive description of children with dengue shock syndrome (DSS) based on a large cohort of children admitted to a single institution with established DSS. The description provides a solid basis for further research in order to achieve a better understanding of the disease. The results demonstrate that case fatality in children with established DSS is very low if diagnosis is prompt, and the patient is immediately admitted to an intensive care unit and carefully managed by an experienced team of clinicians and nurses. Clinicians working in other settings, especially those newly exposed to this disease, could use this experience to improve outcome for their patients.

This thesis also identified several risk factors a) for profound DSS amongst children with DSS, and b) for the development of DSS amongst children hospitalized with dengue. These findings not only provide empirical evidence for experienced clinicians, who may already recognize these factors in their clinical practice, but also provide useful prognostic guidance for clinicians less familiar with the disease. As most of the risk factors identified are readily available in clinical practice, these findings could be very useful for clinicians in regions where resources are limited.

A major contribution of the present thesis is development of the two prognostic models for profound DSS in children with DSS and for progression to DSS in children hospitalized with dengue. A simple score chart was derived from the prediction model for profound

DSS which can be applied in clinical practice, for example to prioritise patient triage, and in research, for example to identify the target population for studies evaluating new interventions for DSS.

In addition, the case study in chapter 6 of this thesis provides preliminary evidence that a patient's current platelet count is a better predictor of DSS than the enrolment value and that dynamic prediction modelling can improve prognostic modelling in dengue.

7.1.2 Statistical contributions

Even though guidelines and standard recommendation for the development of prediction models based on baseline covariates are available (Harrell, 2001; Steyerberg, 2010), explicitly developing a prognostic model in a specific disease still poses challenges. This thesis provides case studies that illustrate how to develop prognostic models for dengue, a task that required special considerations regarding the choice of the outcome and covariates of interest, treatment of missing data, and the potential relevance of dynamic predictions. These case studies are useful for researchers interested in the topic but unfamiliar with prognostic modelling techniques or dengue.

Dynamic prediction models allow the usage of accruing longitudinal information to update predictions and are the topic of active ongoing statistical research. This thesis provides the first systematic comparison of acute and chronic diseases with respect to dynamic predictive modelling. Several differences between the two settings were identified regarding the choice of the time origin, prediction horizon, and outcome of interest; the frequency and regularity of repeated measurement; the expected relationship between outcome and time-dependent covariates; the possibility of competing risks; and the importance of simple and rapid prognostic algorithms. These differences suggest that conditional models which are simpler to develop and interpret than joint models might be preferable in the acute setting. This recommendation was supported by a case study which comparing different approaches to integrating daily platelet counts into a dynamic prediction model for the development of DSS.

7.2 Suggestions for future research

Findings from this thesis suggest further research to investigate the role of risk factors, such as platelet count, haematocrit level and gender, on the pathogenesis of the disease. An interesting finding is that platelet count and haematocrit levels might have different relevance at different stages of the disease: platelet count is an important risk factor for developing DSS in the early stage whereas once a patient has progressed to DSS, haematocrit is more important as an indicator of further progression of the disease. Furthermore, changes in platelet counts over time also relate to changes in the likelihood of developing DSS. As the main underlying pathophysiological abnormality in DSS is plasma leakage (Simmons et al., 2012a), these findings suggest a possible role for platelets in the induction of plasma leakage, a phenomenon supported by the recent work by Hottz et al. (2013). Haematocrit levels, by constrast, are likely to reflect the extent of plasma leakage but not to be involved at a mechanistic level. Further research is required to clearly determine the roles of these factors in the pathophysiology of the disease. In addition, further research is required to confirm the role of gender with respect to the risk of progression to DSS, health-seeking behavior, and the observed interaction between gender and haemodynamic parameters on the risk of progression from DSS to profound DSS.

The simple score chart for prediction of profound DSS developed in chapter 4 has the potential to be a valuable prediction tool for clinicians. However, this score chart was based on a prediction model using data from a single hospital only with moderate performance. Continuing research is required to further assess this score in clinical practice. This includes independent validation studies to assess the performance of the score chart in other settings and subsequent studies to assess the impact of score-chart guided management of DSS patients on outcomes and costs, ideally in a comparative trial (Moons et al., 2009).

The case study in chapter 6 suggests the value of dynamic models for predicting DSS based on longitudinal data. However, this case study was based on a dataset with a relatively low number of DSS events and lacked longitudinal data for laboratory markers and signs and symptoms other than the platelet count. Hence, this case study contributes only as proof-of-concept analysis. A large ongoing prospective multi-centre study within

the International Research Consortium on Dengue Risk Assessment, Management, and Surveillance (Jaenisch et al., 2013) is currently collecting extensive longitudinal data on a large number of dengue patients. The resulting dataset should provide an excellent opportunity to develop a powerful dynamic prediction model using the approaches outlined in this thesis.

References

Aalen OO, Borgan Or, and Gjessing HkK (2008). *Survival and Event History Analysis*. Statistics for Biology and Health. Springer New York, New York, NY.

Alexander N, Balmaseda A, Coelho ICB, Dimaano E, Hien TT, et al. (2011). Multicentre prospective study on dengue classification in four South-east Asian and three Latin American countries. *Tropical Medicine and International Health*, 16(8):936–48.

Almas A, Parkash O, and Akhter J (2010). Clinical factors associated with mortality in dengue infection at a tertiary care center. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 41(2):333–40.

Altman DG and De Stavola BL (1994). Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates. *Statistics in Medicine*, 13(4):301–41.

Anders KL, Nguyet NM, Chau NVV, Hung NT, Thuy TT, et al. (2011). Epidemiological factors associated with dengue shock syndrome and mortality in hospitalized dengue patients in Ho Chi Minh City, Vietnam. *The American Journal of Tropical Medicine and Hygiene*, 84(1):127–34.

Andrinopoulou ER, Rizopoulos D, Takkenberg JJM, and Lesaffre E (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33:3167–78.

Appanna R, Wang SM, Ponnampalavanar Sa, Lum LCS, and Sekaran SD (2012). Cytokine factors present in dengue patient sera induces alterations of junctional proteins in human endothelial cells. *The American Journal of Tropical Medicine and Hygiene*, 87(5):936–42.

Bandyopadhyay S, Lum LCS, and Kroeger A (2006). Classifying dengue: a review of the difficulties in using the WHO case classification for dengue haemorrhagic fever. *Tropical Medicine and International Health*, 11(8):1238–55.

Barnes WJ and Rosen L (1974). Fatal hemorrhagic disease and shock associated with primary dengue infection on a Pacific island. *The American Journal of Tropical Medicine and Hygiene*, 23(3):495–506.

Barniol J, Gaczkowski R, Barbato EV, Da Cunha RV, Salgado D, et al. (2011). Usefulness and applicability of the revised dengue case classification by disease: multi-centre study in 18 countries. *BMC Infectious Diseases*, 11:106.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, et al. (2013). The global distribution and burden of dengue. *Nature*, 496(7446):504–7.

Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues JF, et al. (2014). Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, Preprint:1–20.

Bland JM and Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 1(8476):307–10.

Branco MDRFC, Luna EJDA, Braga Júnior LL, Oliveira RVBD, Rios LTM, et al. (2014). Risk factors associated with death in Brazilian children with severe dengue: a case-control study. *Clinics*, 69(1):55–60.

Brasier AR, Garcia J, Wiktorowicz JE, Spratt HM, Comach G, et al. (2012a). Discovery proteomics and nonparametric modeling pipeline in the development of a candidate biomarker panel for dengue hemorrhagic fever. *Clinical and Translational Science*, 5(1):8–20.

Brasier AR, Ju H, Garcia J, Spratt HM, Victor SS, et al. (2012b). A three-component biomarker panel for prediction of dengue hemorrhagic fever. *The American Journal of Tropical Medicine and Hygiene*, 86(2):341–8.

Bunnag T and Kalayanarooj S (2011). Dengue shock syndrome at the emergency room of Queen Sirikit National Institute of Child Health, Bangkok, Thailand. *Journal of the Medical Association of Thailand*, 94 Suppl 3:S57–63.

Butthep P, Chunhakan S, Yoksan S, Tangnararatchakit K, and Chuansumrit A (2012). Alteration of cytokines and chemokines during febrile episodes associated with endothelial cell damage and plasma leakage in dengue hemorrhagic fever. *The Pediatric Infectious Disease Journal*, 31(12):e232–8.

Capeding MR, Tran NH, Hadinegoro SRS, Ismail HIHM, Chotpitayasunondh T, et al. (2014). Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. *The Lancet*, 384(9951):1358–65.

Cardosa MJ, Wang SM, Sum MSH, and Tio PH (2002). Antibodies against prM protein distinguish between previous infection with dengue and Japanese encephalitis viruses. *BMC Microbiology*, 2:9.

Chau TNB, Quyen NTH, Thuy TT, Tuan NM, Hoang DM, et al. (2008). Dengue in Vietnamese infants–results of infection-enhancement assays correlate with age-related disease epidemiology, and cellular immune responses correlate with disease severity. *The Journal of Infectious Diseases*, 198(4):516–24.

Christensen E, Altman DG, Neuberger J, De Stavola BL, Tygstrup N, et al. (1993). Updating prognosis in primary biliary cirrhosis using a time-dependent Cox regression model. PBC1 and PBC2 trial groups. *Gastroenterology*, 105(6):1865–76.

Chuansumrit A, Puripokai C, Butthep P, Wongtiraporn W, Sasanakul W, et al. (2010). Laboratory predictors of dengue shock syndrome during the febrile stage. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 41(2):326–32.

Coller BAG, Barrett ADT, and Thomas SJ (2011). The development of Dengue vaccines. Introduction. *Vaccine*, 29(42):7219–20.

Collett D (2003). Time-dependent variables. In *Modelling Survival Data in Medical Research*, chapter 8, pages 251–72. Chapman & Hall/CRC, 2nd edition.

Cordeiro MT, Braga-Neto U, Nogueira RMR, and Marques ETa (2009). Reliable classifier to differentiate primary and secondary acute dengue infection based on IgG ELISA. *PLoS One*, 4(4):e4945.

Cox D (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, 34(2):187–220.

Cummings DaT, Iamsirithaworn S, Lessler JT, McDermott A, Prasanthong R, et al. (2009). The impact of the demographic transition on dengue in Thailand: insights from a statistical analysis and mathematical modeling. *PLoS medicine*, 6(9):e1000139.

Cuong HQ, Hien NT, Duong TN, Phong TV, Cam NN, et al. (2011). Quantifying the emergence of dengue in Hanoi, Vietnam: 1998-2009. *PLoS Neglected Tropical Diseases*, 5(9):e1322.

Cuong HQ, Vu NT, Cazelles B, and Boni MF (2013). Spatiotemporal Dynamics of Dengue Epidemics, Southern Vietnam. *Emerging Infectious Diseases*, 19(6):945–53.

Cupples LA, D'Agostino RB, Anderson K, and Kannel WB (1988). Comparison of baseline and repeated measure covariate techniques in the Framingham heart study. *Statistics in Medicine*, 7:205–18.

Da Costa VG, Marques-Silva AC, and Moreli ML (2014). A meta-analysis of the diagnostic accuracy of two commercial NS1 antigen ELISA tests for early dengue virus detection. *PLoS One*, 9(4):e94655.

Dawes RM, Faust D, and Meehl PE (1989). Clinical versus actuarial judgment. *Science*, 243(4899):1668–74.

De Souza VAUF, Fernandes S, Araújo ES, Tateno AF, Oliveira OMNPF, et al. (2004). Use of an immunoglobulin G avidity test to discriminate between primary and secondary dengue virus infections. *Journal of Clinical Microbiology*, 42(4):1782–4.

Dewi L and Nurfitri E (2012). Pediatric logistic organ dysfunction score as a predictive tool of dengue shock syndrome outcomes. *Paediatrica Indonesiana*, 52(2):72–7.

Diggle PJ, Heagerty PJ, Liang KY, and Zeger SL (2002). *Analysis of longitudinal data*. Oxford University Press, Great Britain, 2nd edition.

Dinh The T, Le Thi Thu T, Nguyen Minh D, Tran Van N, Tran Tinh H, et al. (2012). Clinical features of dengue in a large vietnamese cohort: intrinsically lower platelet counts and greater risk for bleeding in adults than children. *PLoS Neglected Tropical Diseases*, 6(6):e1679.

Duangchinda T, Dejnirattisai W, Vasanawathana S, Limpitikul W, Tangthawornchaikul N, et al. (2010). Immunodominant T-cell responses to dengue virus NS3 are associated with DHF. *Proceedings of the National Academy of Sciences of the United States of America*, 107(39):1–6.

Endy TP (2002). Spatial and Temporal Circulation of Dengue Virus Serotypes: A Prospective Study of Primary School Children in Kamphaeng Phet, Thailand. *American Journal of Epidemiology*, 156(1):52–9.

Faisal T, Taib MN, and Ibrahim F (2012). Neural network diagnostic system for dengue patients risk classification. *Journal of Medical Systems*, 36(2):661–76.

Farrar JJ, Hien TT, Horstick O, Hung NT, Jaenisch T, et al. (2013). Dogma in classifying dengue disease. *The American Journal of Tropical Medicine and Hygiene*, 89(2):198–201.

Faucett CL and Thomas DC (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–85.

Figueiredo MAa, Rodrigues LC, Barreto ML, Lima JWO, Costa MCN, et al. (2010). Allergies and diabetes as risk factors for dengue hemorrhagic fever: results of a case control study. *PLoS Neglected Tropical Diseases*, 4(6):e699.

Fisher LD and Lin DY (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, 20(6):145–57.

Fitzmaurice GM, Laird NM, and Ware JH (2004). *Applied Longitudinal Analysis*. John Wiley & Sons, 1st edition.

Friedman J, Hastie T, and Tibshirani R (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22. URL `http://www.jstatsoft.org/v33/i01/`.

Gamble J, Bethell D, Day NP, Loc PP, Phu NH, et al. (2000). Age-related changes in microvascular permeability: a significant factor in the susceptibility of children to shock? *Clinical Science*, 98(2):211–6.

Gibbons RV and Vaughn D (2002). Dengue: an escalating problem. *British Medical Journal*, 324(7353):1563–6.

Giraldo D, Sant'Anna C, Périssé ARS, March MDFP, Souza AP, et al. (2011). Characteristics of children hospitalized with dengue fever in an outbreak in Rio de Janeiro, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 105(10):601–3.

Gjenero-Margan I, Aleraj B, Krajcar D, and al E (2011). Autochthonous dengue fever in Croatia, August-September 2010. *Euro Surveillance*, 16(9):19805.

Gomes ALV, Wee LJK, Khan AM, Gil LHVG, Marques ETa, et al. (2010). Classification of dengue fever patients based on gene expression data using support vector machines. *PLoS One*, 5(6):e11267.

Graf E, Schmoor C, Sauerbrei W, and Schumacher M (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18:2529–45.

Greg Ridgeway with contributions from others (2014). *gbm: Generalized Boosted Regression Models*. URL `https://github.com/harrysouthworth/gbm`. R package version 2.1-06.

Grove WM, Zald DH, Lebow BS, Snitz BE, and Nelson C (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1):19–30.

Gupta V, Yadav TP, Pandey RM, Singh A, Gupta M, et al. (2011). Risk factors of dengue shock syndrome in children. *Journal of Tropical Pediatrics*, 57(6):451–6.

Guzman MG, Alvarez M, and Halstead SB (2013). Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: an historical perspective and role of antibody-dependent enhancement of infection. *Archives of Virology*, 158(7):1445–59.

Guzman MG and Harris E (2014). Dengue. *The Lancet*, 6736(14):1–13.

Guzman MG, Kouri G, Bravo J, Valdes L, Vazquez S, et al. (2002). Effect of age on outcome of secondary dengue 2 infections. *International Journal of Infectious Diseases*, 6(2):118–24.

Hadinegoro SRS (2012). The revised WHO dengue case classification: does the system need to be modified? *Paediatrics and International Child Health*, 32 Suppl 1:33–8.

Halekoh U and Hø jsgaard Sr (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2):1–11.

Halstead SB (2013). Dengue: the syndromic basis to pathogenesis research. Inutility of the 2009 WHO case definition. *The American Journal of Tropical Medicine and Hygiene*, 88(2):212–5.

Halstead SB, Nimmannitya S, and Cohen SN (1970). Observations related to pathogenesis of dengue hemorrhagic fever. IV. Relation of disease severity to antibody response and virus recovered. *The Yale Journal of Biology and Medicine*, 42(5):311–28.

Halstead SB and O'Rourke EJ (1977). Antibody-enhanced dengue virus infection in primate leukocytes. *Nature*, 265(5596):739–41.

Halstead SB, Voulgaropoulos E, and Tien NH (1965). Dengue hemorrhagic fever in South Vietnam: report of the 1963 outbreak. *The American Journal of Tropical Medicine and Hygiene*, 14(5):819–30.

Hammond SN, Balmaseda A, Perez L, Tellez Y, Saborio SI, et al. (2005). Differences in dengue severity in infants, children, and adults in a 3-year hospital-based study in Nicaragua. *American Journal of Tropical Medicine and Hygiene*, 73(6):1063–70.

Hang VT, Nguyet NM, Trung DT, Tricou V, Yoksan S, et al. (2009). Diagnostic accuracy of NS1 ELISA and lateral flow rapid tests for dengue sensitivity, specificity and relationship to viraemia and antibody responses. *PLoS Neglected Tropical Diseases*, 3(1):e360.

Harrell FE (2001). *Regression Modeling Strategies: with applications to linear models, logistic regression, and survival analysis*. Springer-Verlag, New York.

Harrell FE and with contributions from Charles Dupont and many others (2014). Hmisc. URL `http://cran.r-project.org/package=Hmisc`.

Hartmann O, Schuetz P, Albrich WC, Anker SD, Mueller B, et al. (2012). Time-dependent Cox regression: serial measurement of the cardiovascular biomarker proadrenomedullin improves survival prediction in patients with lower respiratory tract infection. *International Journal of Cardiology*, 161(3):166–73.

Hastie TJ, Tibshirani R, and Friedman J (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2nd edition.

Hemingway H, Croft P, Perel P, Hayden JA, Abrams KR, et al. (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *British Medical Journal*, 346(feb05 1):e5595.

Holmes E and Twiddy S (2003). The origin, emergence and evolutionary genetics of dengue virus. *Infection, Genetics and Evolution*, 3(1):19–28.

Horstick O, Farrar J, Lum L, Martinez E, San Martin JL, et al. (2012). Reviewing the development, evidence base, and application of the revised dengue case classification. *Pathogens and Global Health*, 106(2):94–101.

Hottz ED, Lopes JF, Freitas C, Valls-de Souza R, Oliveira MF, et al. (2013). Platelets mediate increased endothelium permeability in dengue through NLRP3-inflammasome activation. *Blood*, 122(20):3405–14.

Hughes MD, Raskino CL, Pocock SJ, Biagini MR, and Burroughs AK (1992). Prediction of short-term survival with an application in primary biliary cirrhosis. *Statistics in Medicine*, 11(13):1731–45.

Huy NT, Thao NT, Ha TT, Lan NT, Nga PT, et al. (2013a). Development of clinical decision rules to predict recurrent shock in dengue. *Critical Care*, 17(6):R280.

Huy NT, Van Giang T, Thuy DHD, Kikuchi M, Hien TT, et al. (2013b). Factors associated with dengue shock syndrome: a systematic review and meta-analysis. *PLoS Neglected Tropical Diseases*, 7(9):e2412.

Huynh TT, Nguyen VC, and Wills B (2013). Resuscitation fluids. *The New England Journal of Medicine*, 369(25):2461.

Ibrahim F, Faisal T, Salim MIM, and Taib MN (2010). Non-invasive diagnosis of risk in dengue patients using bioelectrical impedance analysis and artificial neural network. *Medical and Biological Engineering and Computing*, 48(11):1141–8.

Ibrahim F, Taib MN, Abas WABW, Guan CC, and Sulaiman S (2005). A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN). *Computer Methods and Programs in Biomedicine*, 79(3):273–81.

Innis BL, Nisalak A, Nimmannitya S, Kusalerdchariya S, Chongswasdi V, et al. (1989). An enzyme-linked immunosorbent assay to characterize dengue infections where dengue and Japanese encephalitis co-circulate. *The American Journal of Tropical Medicine and Hygiene*, 40(4):418–27.

Iskandar HR, Mulyo D, Agnes P, and Suryatin Y (2008). Comparison of Pediatric Logistic Organ Dysfunction (Pelod) Score and Pediatric Risk of Mortality (Prism) Iii As a Mortality Predictor in Patients With Dengue Shock Syndrome. *Pediatrics*, 121(Supplement):S129.

Jaenisch T, Sakuntabhai A, and Wilder-Smith A (2013). Dengue Research Funded by the European Commission-Scientific Strategies of Three European Dengue Research Consortia. *PLoS Neglected Tropical Diseases*, 7(12):e2320.

Ju H and Brasier AR (2013). Variable selection methods for developing a biomarker panel for prediction of dengue hemorrhagic fever. *BMC Research Notes*, 6(1):365.

Kalayanarooj S (2011). Dengue classification: current WHO vs. the newly suggested classification for better clinical application? *Journal of the Medical Association of Thailand*, 94 Suppl 3:S74–84.

Karp I, Abrahamowicz M, Bartlett G, and Pilote L (2004). Updated risk factor values and the ability of the multivariable risk score to predict coronary heart disease. *American Journal of Epidemiology*, 160(7):707–16.

Khor CC, Chau TNB, Pang J, Davila S, Long HT, et al. (2011). Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. *Nature Genetics*, 43(11):1139–41.

Kliks SC, Nimmanitya S, Nisalak A, and Burke DS (1988). Evidence that maternal dengue antibodies are important in the development of dengue hemorrhagic fever in infants. *The American Journal of Tropical Medicine and Hygiene*, 38(2):411–9.

Koenker R (2005). *Quantile Regression*. Cambridge University Press.

Kuno G, Gómez I, and Gubler DJ (1991). An ELISA procedure for the diagnosis of dengue infections. *Journal of Virological Methods*, 33(1-2):101–13.

Kuo MC, Lu PL, Chang JM, Lin MY, Tsai JJ, et al. (2008). Impact of renal failure on the outcome of dengue viral infection. *Clinical Journal of the American Society of Nephrology*, 3(5):1350–6.

La Ruche G, Souarès Y, Armengaud A, Peloux-Petiot F, Delaunay P, et al. (2010). First two autochthonous dengue virus infections in metropolitan France, September 2010. *Euro Surveillance*, 15(39):19676.

Lanciotti RS, Calisher CH, Gubler DJ, Chang GJ, and Vorndam aV (1992). Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. *Journal of Clinical Microbiology*, 30(3):545–51.

Lawrence Gould a, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, et al. (2014). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine*, Preprint.

Lee VJ, Lye DC, Sun Y, and Leo YS (2009). Decision tree algorithm in deciding hospitalization for adult patients with dengue haemorrhagic fever in Singapore. *Tropical Medicine and International Health*, 14(9):1154–9.

Lee VJ, Lye DCB, Sun Y, Fernandez G, Ong A, et al. (2008). Predictive value of simple clinical and laboratory variables for dengue hemorrhagic fever in adults. *Journal of Clinical Virology*, 42(1):34–9.

Leitmeyer KC, Vaughn DW, Watts DM, Salas R, De Chacon IV, et al. (1999). Dengue virus structural differences that correlate with pathogenesis. *Journal of virology*, 73(6):4738–47.

Lemeshow S, Klar J, Teres D, Avrunin JS, Gehlbach SH, et al. (1994). Mortality probability models for patients in the intensive care unit for 48 or 72 hours. *Critical Care Medicine*, 22(9):1351–8.

Lemeshow S, Teres D, Avrunin JS, and Gage RW (1988). Refining intensive care unit outcome prediction by using changing probabilities of mortality. *Critical Care Medicine*, 16(5):470–7.

Liestø l K and Andersen PK (2002). Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. *Statistics in Medicine*, 21(23):3701–14.

Low JG, Sung C, Wijaya L, Wei Y, Rathore APS, et al. (2014). Efficacy and safety of celgosivir in patients with dengue fever (CELADEN): a phase 1b, randomised, double-blind, placebo-controlled, proof-of-concept trial. *The Lancet Infectious Diseases*, 14(8):706–15.

Maitland K, Kiguli S, and Opoka R (2011). Mortality after fluid bolus in African children with severe infection. *The New England Journal of Medicine*, 364(26):2483–95.

Marón GM, Clará AW, Diddle JW, Pleités EB, Miller L, et al. (2010). Association between nutritional status and severity of dengue infection in children in El Salvador. *The American Journal of Tropical Medicine and Hygiene*, 82(2):324–9.

Marón GM, Escobar GA, Hidalgo EM, Clara AW, Minniear TD, et al. (2011). Characterization of dengue shock syndrome in pediatric patients in El Salvador. *The Pediatric Infectious Disease Journal*, 30(5):449–50.

Matheus S, Deparis X, Labeau B, Lelarge J, Morvan J, et al. (2005). Discrimination between primary and secondary dengue virus infection by an immunoglobulin G avidity test using a single acute-phase serum sample. *Journal of Clinical Microbiology*, 43(6):2793–7.

Mayo Clinic (2015). The Updated Natural History Model for Primary Biliary Cirrhosis - For Medical Professionals - Mayo Clinic. URL `http://www.mayoclinic.org/medical-professionals/model-end-stage-liver-disease/updated-natural-history-model-for-primary-biliary-cirrhosis`.

Meehl PE (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press, Minneapolis, MN, US.

Mena Lora AJ, Fernandez J, Morales A, Soto Y, Feris-Iglesias J, et al. (2014). Disease severity and mortality caused by dengue in a Dominican pediatric population. *The American Journal of Tropical Medicine and Hygiene*, 90(1):169–72.

Meng XL and Rubin DB (1992). Performing Likelihood Ratio Tests with Multiply-Imputed Data Sets. *Biometrika*, 79(1):103–11.

Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, et al. (2014). Global spread of dengue virus types: mapping the 70 year history. *Trends in Microbiology*, 22(3):138–46.

Miettinen OS (2011). *Up from Clinical Epidemiology & EBM*. Springer Netherlands, Dordrecht.

Mongkolsapaya J, Dejnirattisai W, Xu Xn, Vasanawathana S, Tangthawornchaikul N, et al. (2003). Original antigenic sin and apoptosis in the pathogenesis of dengue hemorrhagic fever. *Nature Medicine*, 9(7):921–7.

Moons KGM, Altman DG, Vergouwe Y, and Royston P (2009). Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *British Medical Journal*, 338:b606.

Moxon C and Wills B (2008). Management of severe dengue in children. *Advances in Experimental Medicine and Biology*, 609:131–44.

Murtaugh Pa, Dickson ER, Van Dam GM, Malinchoc M, Grambsch PM, et al. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology*, 20(1):126–34.

Myburgh Ja and Mythen MG (2013). Resuscitation fluids. *The New England Journal of Medicine*, 369(13):1243–51.

Nguyen NM, Tran CNB, Phung LK, Duong KTH, Huynh HLA, et al. (2013). A randomized, double-blind placebo controlled trial of balapiravir, a polymerase inhibitor, in Adult dengue patients. *The Journal of Infectious Diseases*, 207(9):1442–50.

Nisalak A, Endy TP, Nimmannitya S, Kalayanarooj S, Thisayakorn U, et al. (2003). Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. *The American Journal of Tropical Medicine and Hygiene*, 68(2):191–202.

Nogueira R, Miagostovich M, Da Cunha RV, Zagne S, Gomes F, et al. (1999). Fatal primary dengue infections in Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(4):418.

Pang J, Salim A, Lee VJ, Hibberd ML, Chia KS, et al. (2012). Diabetes with hypertension as risk factors for adult dengue hemorrhagic fever in a predominantly dengue serotype 2 epidemic: a case control study. *PLoS Neglected Tropical Diseases*, 6(5):e1641.

Pawitan Ja (2011). Dengue virus infection: predictors for severe dengue. *Acta Medica Indonesiana*, 43(2):129–35.

Peeling RW, Artsob H, Pelegrino JL, Buchy P, Cardosa MJ, et al. (2010). Evaluation of diagnostic tests: dengue. *Nature Reviews Microbiology*, 8(12):S30–S37.

Pepe M and Couper D (1997). Modeling partly conditional means with longitudinal data. *Journal of the American Statistical Association*, 92(439):991–8.

Pepe M, Heagerty PJ, and Whitaker R (1999). Prediction using partly conditional time-varying coefficients regression models. *Biometrics*, 55(3):944–50.

Phuong CXT, Nhan NT, Kneen R, Thuy PTT, van Thien C, et al. (2004). Clinical diagnosis and assessment of severity of confirmed dengue infections in Vietnamese children: is the world health organization classification system helpful? *The American Journal of Tropical Medicine and Hygiene*, 70(2):172–9.

Phuong HL, Vries PJD, Thai KTD, Nga TTT, Hung LQ, et al. (2006). Dengue Virus Infections in Viet Nam: Tip of the Iceberg. *Dengue Bulletin*, 30:15–25.

Pollack MM, Ruttimann UE, and Getson PR (1988). Pediatric risk of mortality (PRISM) score. *Critical Care Medicine*, 16(11):1110–6.

Pongpan S, Wisitwong A, Tawichasri C, Patumanond J, and Namwongprom S (2013). Development of dengue infection severity score. *ISRN Pediatrics*, 2013:1–6.

Potts JA, Gibbons RV, Rothman AL, Srikiatkhachorn A, Thomas SJ, et al. (2010a). Prediction of dengue disease severity among pediatric Thai patients using early clinical laboratory indicators. *PLoS Neglected Tropical Diseases*, 4(8):e769.

Potts JA and Rothman AL (2008). Clinical and laboratory features that distinguish dengue from other febrile illnesses in endemic populations. *Tropical Medicine and International Health*, 13(11):1328–40.

Potts JA, Thomas SJ, Srikiatkhachorn A, Supradish Po, Li W, et al. (2010b). Classification of dengue illness based on readily available laboratory data. *The American Journal of Tropical Medicine and Hygiene*, 83(4):781–8.

Proust-Lima C, Séne M, Taylor JMG, and Jacqmin-Gadda H (2014). Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research*, 23(1):74–90.

Proust-Lima C and Taylor JMG (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics*, 10(3):535–49.

Pun SB (2011). Dengue: an emerging disease in Nepal. *Journal of the Nepal Medical Association*, 51(184):203–8.

Quang Ha D, Ha D, Tien N, and Huong V (2000). Dengue epidemic in southern Vietnam, 1998. *Emerging Infectious Diseases*, 6(4):1998–2001.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.r-project.org/`.

Ranjit S, Kissoon N, and Jayakumar I (2005). Aggressive management of dengue shock syndrome may decrease mortality rate: a suggested protocol. *Pediatric Critical Care Medicine*, 6(4):412–9.

Recker M, Blyuss KB, Simmons CP, Hien TT, Wills B, et al. (2009). Immunological serotype interactions and their effect on the epidemiological pattern of dengue. *Proceedings of the Royal Society B*, 276(1667):2541–8.

Reilly BM and Evans AT (2006). Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Annals of Internal Medicine*, 144(3):201–9.

Rico-Hesse R, Harrison LM, Salas RA, Tovar D, Nisalak A, et al. (1997). Origins of Dengue Type 2 Viruses Associated with Increased Pathogenicity in the Americas. *Virology*, 230(2):244–51.

Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams KR, et al. (2013). Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLoS Medicine*, 10(2):e1001380.

Rizopoulos D (2010). JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. *Journal of Statistical Software*, 35(9):1–33.

Rizopoulos D (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–29.

Rizopoulos D (2012). *Joint Models for Longitudinal and Time-to-event Data*. Chapman and Hall/CRC.

Rizopoulos D and Ghosh P (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30:1366–80.

Rizopoulos D, Verbeke G, and Molenberghs G (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74.

Rué M, Quintana S, Alvarez M, and Artigas A (2001). Daily assessment of severity of illness and mortality prediction for individual patients. *Critical Care Medicine*, 29(1):45–50.

Ruttimann UE and Pollack MM (1991). Objective assessment of changing mortality risks in pediatric intensive care unit patients. *Critical Care Medicine*, 19(4):474–83.

Sabchareon A, Wallace D, Sirivichayakul C, Limkittikul K, Chanthavanich P, et al. (2012). Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. *The Lancet*, 6736(12):1–9.

Sakuntabhai A, Turbpaiboon C, Casadémont I, Chuansumrit A, Lowhnoo T, et al. (2005). A variant in the CD209 promoter is associated with severity of dengue disease. *Nature Genetics*, 37(5):507–13.

Salje H, Lessler J, Endy TP, Curriero FC, Gibbons RV, et al. (2012). Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. *Proceedings of the National Academy of Sciences of the United States of America*, 109(24):9535–8.

Sangkawibha N, Rojanasuphot S, Ahandrik S, Viriyapongse S, Jatanasen S, et al. (1984). Risk factors in dengue shock syndrome: a prospective epidemiologic study in Rayong, Thailand. I. The 1980 outbreak. *American Journal of Epidemiology*, 120(5):653–69.

Schoop R, Graf E, and Schumacher M (2008). Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics*, 64(2):603–10.

Schoop R, Schumacher M, and Graf E (2011). Measures of prediction error for survival data with longitudinal covariates. *Biometrical Journal*, 53(2):275–93.

Scott RM, Nimmannitya S, Bancroft WH, and Mansuwan P (1976). Shock syndrome in primary dengue infections. *The American Journal of Tropical Medicine and Hygiene*, 25(6):866–74.

Shann F, Pearson G, Slater A, and Wilkinson K (1997). Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive Care Medicine*, 23(2):201–7.

Shepard DS, Undurraga Ea, and Halasa Ya (2013). Economic and disease burden of dengue in Southeast Asia. *PLoS Neglected Tropical Diseases*, 7(2):e2055.

Shu Py, Chang Sf, Kuo Yc, Yueh Yy, Chien Lj, et al. (2003a). Development of group- and serotype-specific one-step SYBR green I-based real-time reverse transcription-PCR assay for dengue virus. *Journal of Clinical Microbiology*, 41(6):2408–16.

Shu PY, Chen LK, Chang SF, Yueh YY, Chow L, et al. (2003b). Comparison of capture immunoglobulin M (IgM) and IgG enzyme-linked immunosorbent assay (ELISA) and nonstructural protein NS1 serotype-specific IgG ELISA for differentiation of primary and secondary dengue virus infections. *Clinical and Diagnostic Laboratory Immunology*, 10(4):622–30.

Sierra BDLC, Kourí G, and Guzman MG (2007). Race: A risk factor for dengue hemorrhagic fever. *Archives of Virology*, 152(3):533–42.

Simmons CP, Farrar JJ, Nguyen vVC, and Wills B (2012a). Dengue. *The New England Journal of Medicine*, 366(15):1423–32.

Simmons CP, Wolbers M, Nguyen MN, Whitehorn J, Shi PY, et al. (2012b). Therapeutics for dengue: recommendations for design and conduct of early-phase clinical trials. *PLoS Neglected Tropical Diseases*, 6(9):e1752.

Singer JD and Willett JB (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press.

Snow GE, Haaland B, Ooi EE, and Gubler DJ (2014). Research on Dengue during World War II Revisited. *The American Journal of Tropical Medicine and Hygiene*, 91(6):1203–17.

Sox HC, Higgins MC, and Owens DK (2013). *Medical Decision Making*. John Wiley & Sons, Ltd, Chichester, UK, 2nd edition.

Srikiatkhachorn A and Green S (2010). Markers of dengue disease severity. *Current Topics in Microbiology and Immunology*, 338:67–82.

Srikiatkhachorn A, Krautrachue A, Ratanaprakarn W, Wongtapradit L, Nithipanya N, et al. (2007). Natural history of plasma leakage in dengue hemorrhagic fever: a serial ultrasonographic study. *The Pediatric Infectious Disease Journal*, 26(4):283–90; discussion 291–2.

Srikiatkhachorn A, Rothman AL, Gibbons RV, Sittisombut N, Malasit P, et al. (2011). Dengue–how best to classify it. *Clinical Infectious Diseases*, 53(6):563–7.

Sterne JaC, White IR, Carlin JB, Spratt M, Royston P, et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal*, 338:b2393.

Steyerberg EW (2010). *Clinical Prediction Models: a practical approach to development, validation, and updating*. Springer, New York.

Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, et al. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Medicine*, 10(2):e1001381.

Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan Va, Vazquez-Prokopec GM, et al. (2013). House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 110(3):994–9.

Suaya Ja, Shepard DS, Siqueira JaB, Martelli CT, Lum LCS, et al. (2009). Cost of dengue cases in eight countries in the Americas and Asia: a prospective study. *The American Journal of Tropical Medicine and Hygiene*, 80(5):846–55.

Sullivan LM, Massaro JM, and D'Agostino RB (2004). Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Statistics in Medicine*, 23(10):1631–60.

Sweeting MJ and Thompson SG (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–63.

Tam DTH, Ngoc TV, Tien NTH, Kieu NTT, Thuy TTT, et al. (2012a). Effects of short-course oral corticosteroid therapy in early dengue infection in Vietnamese patients: a randomized, placebo-controlled trial. *Clinical Infectious Diseases*, 55(9):1216–24.

Tam PT, Dat NT, Huu LM, Thi XCP, Duc HM, et al. (2012b). High household economic burden caused by hospitalization of patients with severe dengue fever cases in Can Tho province, Vietnam. *The American Journal of Tropical Medicine and Hygiene*, 87(3):554–8.

Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, et al. (2008). Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS Neglected Tropical Diseases*, 2(3):e196.

Tantracheewathorn T and Tantracheewathorn S (2007). Risk factors of dengue shock syndrome in children. *Journal of the Medical Association of Thailand*, 90(2):272–7.

Tee HP, How SH, Jamalludin AR, Safhan MNF, Sapian MM, et al. (2009). Risk factors associated with development of dengue haemorrhagic fever or dengue shock syndrome in adults in Hospital Tengku Ampuan Afzan Kuantan. *Medical Journal of Malaysia*, 64(4):316–20.

Thein TL, Leo YS, Lee VJ, Sun Y, and Lye DC (2011). Validation of probability equation and decision tree in predicting subsequent dengue hemorrhagic fever in adult dengue inpatients in Singapore. *The American Journal of Tropical Medicine and Hygiene*, 85(5):942–5.

Therneau T, Atkinson B, and Ripley B (2014). *rpart: Recursive Partitioning and Regression Trees*. URL `http://CRAN.R-project.org/package=rpart`. R package version 4.1-8.

Tobin J (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26(1):24.

Tricou V, Minh NN, Van TP, Lee SJ, Farrar JJ, et al. (2010a). A randomized controlled trial of chloroquine for the treatment of dengue in Vietnamese adults. *PLoS Neglected Tropical Diseases*, 4(8):e785.

Tricou V, Vu HTT, Quynh NVN, Nguyen CVV, Tran HT, et al. (2010b). Comparison of two dengue NS1 rapid tests for sensitivity, specificity and relationship to viraemia and antibody responses. *BMC Infectious Diseases*, 10:142.

Trung DT and Wills B (2010). Systemic vascular leakage associated with dengue infections - the clinical perspective. *Current Topics in Microbiology and Immunology*, 338:57–66.

Tsai JJ, Chokephaibulkit K, Chen PC, Liu LT, Hsiao HM, et al. (2013). Role of cognitive parameters in dengue hemorrhagic fever and dengue shock syndrome. *Journal of Biomedical Science*, 20:88.

Tsiatis AA and Davidian M (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14:809–34.

Van Buuren S and Groothuis-Oudshoorn K (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3):1–67.

Van Houwelingen HC and Putter H (2012). *Dynamic prediction in clinical survival analysis*. Taylor & Francis Group.

Vaughn DW, Green S, Kalayanarooj S, Innis BL, Nimmannitya S, et al. (1997). Dengue in the early febrile phase: viremia and antibody responses. *The Journal of Infectious Diseases*, 176(2):322–30.

Vaughn DW, Green S, Kalayanarooj S, Innis BL, Nimmannitya S, et al. (2000). Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *The Journal of Infectious Diseases*, 181(1):2–9.

Vaughn DW, Nisalak A, Solomon T, Kalayanarooj S, Nguyen MD, et al. (1999). Rapid serologic diagnosis of dengue virus infection using a commercial capture ELISA that distinguishes primary and secondary infections. *The American Journal of Tropical Medicine and Hygiene*, 60(4):693–8.

Vejbaesya S, Luangtrakool P, Luangtrakool K, Kalayanarooj S, Vaughn DW, et al. (2009). TNF and LTA gene, allele, and extended HLA haplotype associations with severe dengue virus infection in ethnic Thais. *The Journal of Infectious Diseases*, 199(10):1442–8.

Verbeke G and Davidian M (2008). Joint models for longitudinal data. In Verbeke G, Davidian M, Fitzmaurice G, and Molenberghs G, editors, *Longitudinal Data Analysis*, volume 20085746 of *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*, chapter 13, pages 319–26. Chapman and Hall/CRC.

Villar L, Dayan GH, Arredondo-García JL, Rivera DM, Cunha R, et al. (2014). Efficacy of a Tetravalent Dengue Vaccine in Children in Latin America. *New England Journal of Medicine*, Preprint.

Vu TTH, Holmes EC, Duong V, Nguyen TQ, Tran TH, et al. (2010). Emergence of the Asian 1 genotype of dengue virus serotype 2 in viet nam: in vivo fitness advantage and lineage replacement in South-East Asia. *PLoS Neglected Tropical Diseases*, 4(7):e757.

Wagner DP, Knaus WA, Harrell FE, Zimmerman JE, and Watts C (1994). Daily prognostic estimates for critically ill adults in intensive care units: results from a prospective, multicenter, inception cohort analysis. *Critical Care Medicine*, 22(9):1359–72.

White IR, Royston P, and Wood AM (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–99.

Whitehorn J and Simmons CP (2011). The pathogenesis of dengue. *Vaccine*, 29(42):7221–8.

Whitehorn J, Van Vinh Chau N, Truong NT, Tai LTH, Van Hao N, et al. (2012). Lovastatin for adult patients with dengue: protocol for a randomised controlled trial. *Trials*, 13(1):203.

Wichmann O, Hongsiriwon S, Bowonwatanuwong C, Chotivanich K, Sukthana Y, et al. (2004). Risk factors and clinical features associated with severe dengue infection in adults and children during the 2001 epidemic in Chonburi , Thailand. *Tropical Medicine and International Health*, 9(9):1022–9.

Wickham H (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. URL `http://had.co.nz/ggplot2/book`.

Wickham H (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29. URL `http://www.jstatsoft.org/v40/i01/`.

Wickham H and Francois R (2014). *dplyr: A Grammar of Data Manipulation*. URL `http://CRAN.R-project.org/package=dplyr`. R package version 0.3.0.2.

Wilder-Smith A (2014). Dengue vaccines: dawning at last? *The Lancet*, 6736(14):1327–9.

Wills B, Nguyen MD, Ha TL, Dong Thi Hoai T, Tran TNT, et al. (2005). Comparison of three fluid solutions for resuscitation in dengue shock syndrome. *The New England Journal of Medicine*, 353(9):877–89.

Wills B, Oragui EE, Stephens AC, Daramola Oa, Dung NM, et al. (2002). Coagulation abnormalities in dengue hemorrhagic Fever: serial investigations in 167 Vietnamese children with Dengue shock syndrome. *Clinical Infectious Diseases*, 35(3):277–85.

Wills B, Tran VN, Nguyen THV, Truong TTT, Tran TNT, et al. (2009). Hemostatic changes in Vietnamese children with mild dengue correlate with the severity of vascular leakage rather than bleeding. *The American Journal of Tropical Medicine and Hygiene*, 81(4):638–44.

Wolbers M, Koller MT, Witteman JCM, and Steyerberg EW (2009). Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology*, 20(4):555–61.

Wood S (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 1 edition.

Wood SN (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36.

World Health Organization (1997). *Dengue haemorrhagic fever: diagnosis, treatment, prevention and control*. World Health Organization, Geneva, 2nd edition.

World Health Organization (2009). *Dengue Guidelines for Diagnosis, Treatment, Prevention and Control*. World Health Organization, France, 2009 edition.

World Health Organization (2012a). Dengue and severe dengue. URL `http://www.who.int/mediacentre/factsheets/fs117/en/http://www.who.int/mediacentre/factsheets/fs117/en/index.html`.

World Health Organization (2012b). *Global strategy for dengue prevention and control 2012-2020*. World Health Organization.

World Health Organization (2012c). *Handbook for clinical management of dengue*. World Health Organization.

World Health Organization (2012d). Questions and Answers on Dengue Vaccines : Phase IIb study of CYD-TDV.

Wu M and Ware JH (1979). On the Use of Repeated Measurements in Regression Analysis with Dichotomous Responses. *Biometrics*, 35(2):513–21.

Wulfsohn MS and Tsiatis aa (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–9.

Yacoub S, Griffiths A, Chau TTH, Simmons CP, Wills B, et al. (2012). Cardiac function in Vietnamese patients with different dengue severity grades. *Critical Care Medicine*, 40(2):477–83.

Yacoub S, Mongkolsapaya J, and Screaton G (2013). The pathogenesis of dengue. *Current Opinion In Infectious Diseases*, 26(3):284–9.

Zhang H, Li W, Wang J, Peng H, Che X, et al. (2014). NS1-based tests with diagnostic utility for confirming dengue infection: a meta-analysis. *International Journal of Infectious Diseases*, 26:57–66.

Zheng Y and Heagerty PJ (2005). Partly conditional survival models for longitudinal data. *Biometrics*, 61(2):379–91.

Zheng Y and Heagerty PJ (2007). Prospective accuracy for longitudinal markers. *Biometrics*, 63(2):332–41.